An evaluation of the effectiveness of a computer-aided assessment system for mathematics and engineering students

Author: Stephen James Broughton

Supervisors: Dr. Paul Hernandez-Martinez
Prof. Carol L. Robinson

A thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy in the Mathematics Education Centre School of Science

April 5, 2017
Declaration of Authorship

I, Stephen James Broughton, declare that this thesis titled, 'An evaluation of the effectiveness of a computer-aided assessment system for mathematics and engineering students' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:
“Those who have taken upon them to lay down the law of nature as a thing already searched out and understood, whether they have spoken in simple assurance or professional affectation, have therein done philosophy and the sciences great injury.”

Francis Bacon

“Language is only the instrument of science, and words are but the signs of ideas: I wish, however, that the instrument might be less apt to decay, and that signs might be permanent, like the things which they denote.”

Samuel Johnson
UNIVERSITY OF LOUGHBOROUGH

Abstract

School of Science
Mathematics Education Centre

Master of Philosophy

An evaluation of the effectiveness of a computer-aided assessment system for mathematics and engineering students

by Stephen James Broughton

Computer-aided assessment is a means by which to assess many students quickly and efficiently. Its popularity as a tool for assessing mathematics increased substantially in the first decade of this millennium as computing and the Internet became more widely available, and as cohort sizes grew.

This research sought to evaluate one such system that had been used at a higher education institution for over ten years. However, the literature does not offer a clear or detailed framework from which to perform an evaluation of this system.

Using cultural-historical activity theory, and cues from assessment literature, this research presents a model for effective assessment. It provides a framework for judging where an assessment is effective for individuals using the assessment and where it ceases to be effective.

Case study analyses helped to identify where the assessment tool was no longer effective. It identified that students struggled to construct new goals after the summative phase of assessment. It also explained how and why the lecturers had diverse practices.
Acknowledgements

In some ways this has been the hardest section to write — this thesis has been a spectacular journey that has only been made possible by so many people; yet I am expected to limit my gratitude to one page. As you will see from reading this, I am fond of exceeding expectations.

I have dedicated this thesis to my mother, Lesley Broughton, who never seems to say “no” whenever I have an outlandish idea, an extravagant goal, or a need for somewhere to stay. She has not been so much of a rock to me; more solid concrete foundations, four walls, a bed, a radiator, a shoulder to cry on and endless source of washing powder. This would not have been possible without her: she is the best.

In writing this thesis, I also lost members of my family, so I wish to acknowledge their contribution to my life and my path to this point. My uncle, Michael Leeks: he was an ever-loving and fun uncle. He taught me how to ride a bicycle, which I have remembered as a moment of accomplishment and struck a thirst within me for more. My grandmother, Cynthia Moat: she nurtured my musical creativity, which in the long-run was a sound investment for my sanity. My grandfather, Donald Broughton, “Papa”: he is an eternal inspiration, a loving guide, equally as invested in my success as I am. We both aimed high and did not quite achieve what we had set out to, but it will happen.

My friends are extraordinarily important to me, and those that have stuck by me will stay in my heart (and my thesis) forever. I am lucky that they are too numerable to mention all by name, but I pay special thanks to Phil Broadbridge, Nick Dathan, Caleb Emerson, Maria Harbak, Peter Harris, Rob Jones, Simon Llewellyn, James Newell, Racko Scythewing, Mike Vucek and Jack Wilkin, who have all been instrumental in seeing me through to the finish of this thesis.

The Mathematics Education Centre at Loughborough University is an incredible institution. Professionally, academically and socially, they have left an indelible (in a good way) impression on how I have become as a researcher and as a person. I thank especially Professors Tony Croft and Barbara Jaworski in seeing something in me to acquire the funding for me to study here. Having lunch every day with members of the MEC saw me through some of my darkest days; this should not be underestimated and every one of them — past and present — can say they have supported me in this endeavour.

I have since settled at Oxford Brookes University, which has been just as supportive and encouraging as I balanced several commitments as well as this thesis. Professor Rhona Sharpe has been an incredible manager. She has managed to keep me working — and productively, and happily — while I finished my thesis through ill-health. She nurtured
my determination and, as a result, has helped me to succeed when I might have given up. Furthermore, members of my department in the Oxford Centre for Staff and Learning Development have become an excellent support network.

It is customary to laud the contribution of one’s supervisors at this point. However, this task seems impossible. Dr Paul Hernandez-Martinez and Professor Carol Robinson did not just supervise a research student: they were the first people to whom I disclosed my depression and the first to direct me to the support I needed; they were optimists, realists, motivators, counsellors and bearers-of-bad-news at moments when those roles were demanded; they gave expert guidance and directed my enthusiasm towards international conferences, competitions, bids and presentations. It was an extraordinary journey in which they surely had to go above and beyond the remit of their role as supervisors and I am eternally grateful to them.

I am eternally grateful to you all. Thank you.
# Contents

Declaration of Authorship .................................................. i

Abstract .................................................................................. iii

Acknowledgements ................................................................... iv

Contents .................................................................................. vi

List of Figures ........................................................................... ix

List of Tables ............................................................................. xii

Abbreviations ............................................................................ xiii

1 Introduction ............................................................................ 1

2 Literature Review ................................................................... 4  
   2.1 Making assessment effective ............................................. 5  
      2.1.1 Current assessment practice ...................................... 6  
      2.1.2 The role of the lecturer ............................................ 8  
      2.1.3 The role of the student ............................................ 12  
      2.1.4 A definition for effective assessment ..................... 13  
      2.1.5 Problems with formative assessment ..................... 23  
      2.1.6 Developing a definition of effective assessment .... 24  
   2.2 Computer-aided assessment ............................................. 27  
      2.2.1 What is computer-aided assessment? ...................... 28  
      2.2.2 The benefits of computer-aided assessment ........... 29  
      2.2.3 The use of computer-aided assessment in higher education mathematics learning in the United Kingdom ............................................. 33  
      2.2.4 The problems with computer-aided assessment ....... 38  
      2.2.5 Summary .............................................................. 42  
   2.3 Conclusions ....................................................................... 43

3 Theoretical Framework ......................................................... 45  
   3.1 What is activity theory? .................................................... 45
<table>
<thead>
<tr>
<th>Contents</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1 The structure of activity . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>46</td>
</tr>
<tr>
<td>3.1.2 Key ideas within activity theory . . . . . . . . . . . . . . . . . . .</td>
<td>47</td>
</tr>
<tr>
<td>3.2 Detailed examination of aspects of CHAT . . . . . . . . . . . . . . . .</td>
<td>49</td>
</tr>
<tr>
<td>3.2.1 Developing an identity . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>49</td>
</tr>
<tr>
<td>3.2.2 Goals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>53</td>
</tr>
<tr>
<td>3.2.3 The influence of the community . . . . . . . . . . . . . . . . . . . . .</td>
<td>60</td>
</tr>
<tr>
<td>3.2.4 Resolving contradictions . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>65</td>
</tr>
<tr>
<td>3.2.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>67</td>
</tr>
<tr>
<td>3.3 Proposing a model for effective assessment . . . . . . . . . . . . . . .</td>
<td>68</td>
</tr>
<tr>
<td>4 Methodology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>73</td>
</tr>
<tr>
<td>4.1 A case study approach . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>74</td>
</tr>
<tr>
<td>4.2 Choosing methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>76</td>
</tr>
<tr>
<td>4.2.1 Ruling out observations . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>77</td>
</tr>
<tr>
<td>4.3 Questionnaires . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>79</td>
</tr>
<tr>
<td>4.4 Interviews . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>81</td>
</tr>
<tr>
<td>4.4.1 Thematising . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>82</td>
</tr>
<tr>
<td>4.4.2 Designing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>83</td>
</tr>
<tr>
<td>4.4.3 Interviewing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>84</td>
</tr>
<tr>
<td>4.4.4 Transcribing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>86</td>
</tr>
<tr>
<td>4.4.5 Analysing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>87</td>
</tr>
<tr>
<td>4.4.6 Verifying . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>87</td>
</tr>
<tr>
<td>4.4.7 Reporting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>89</td>
</tr>
<tr>
<td>4.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>89</td>
</tr>
<tr>
<td>5 Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>91</td>
</tr>
<tr>
<td>5.1 The student study . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>92</td>
</tr>
<tr>
<td>5.1.1 Student questionnaires . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>92</td>
</tr>
<tr>
<td>5.1.2 Student interviews . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>93</td>
</tr>
<tr>
<td>5.1.3 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>116</td>
</tr>
<tr>
<td>5.2 The lecturer study . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>118</td>
</tr>
<tr>
<td>5.2.1 Lecturer questionnaires . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>119</td>
</tr>
<tr>
<td>5.2.2 Lecturer interviews . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>122</td>
</tr>
<tr>
<td>5.2.3 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>146</td>
</tr>
<tr>
<td>5.3 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>148</td>
</tr>
<tr>
<td>5.3.1 What makes an assessment “effective”? . . . . . . . . . . . . . . . . . .</td>
<td>148</td>
</tr>
<tr>
<td>5.3.2 Under what conditions is the CAA system an effective assessment? . .</td>
<td>149</td>
</tr>
<tr>
<td>5.3.3 What would make the CAA system more effective as an assessment tool?</td>
<td>150</td>
</tr>
<tr>
<td>6 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>153</td>
</tr>
<tr>
<td>6.1 How can an assessment tool be evaluated? . . . . . . . . . . . . . . . . .</td>
<td>153</td>
</tr>
<tr>
<td>6.2 Is the CAA system an effective assessment? . . . . . . . . . . . . . . . .</td>
<td>155</td>
</tr>
<tr>
<td>6.3 How can this model for effective assessment be used elsewhere? . . . . .</td>
<td>156</td>
</tr>
<tr>
<td>A CAA system examples . . . . . . . . . . . . . . . . . . . . . . . . . . . .</td>
<td>158</td>
</tr>
<tr>
<td>Contents</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>B Student questionnaires</td>
<td>160</td>
</tr>
<tr>
<td>B.1 Semester 1 questionnaire</td>
<td>160</td>
</tr>
<tr>
<td>B.2 Semester 2 questionnaire</td>
<td>163</td>
</tr>
<tr>
<td>C Staff questionnaire</td>
<td>167</td>
</tr>
<tr>
<td>D Example of questions used in the student interviews</td>
<td>170</td>
</tr>
<tr>
<td>E Example of questions used in the lecturer interviews</td>
<td>171</td>
</tr>
<tr>
<td>F Student questionnaire analysis</td>
<td>172</td>
</tr>
<tr>
<td>F.1 Cohort attitudes towards mathematics</td>
<td>172</td>
</tr>
<tr>
<td>F.2 The cohort’s approaches to CAA</td>
<td>178</td>
</tr>
<tr>
<td>F.3 Discussion of the student questionnaires</td>
<td>182</td>
</tr>
<tr>
<td>G Student interview case studies</td>
<td>183</td>
</tr>
<tr>
<td>G.1 Alpha</td>
<td>183</td>
</tr>
<tr>
<td>G.2 Beta</td>
<td>188</td>
</tr>
<tr>
<td>G.3 Gamma</td>
<td>193</td>
</tr>
<tr>
<td>G.4 Delta</td>
<td>197</td>
</tr>
<tr>
<td>G.5 Epsilon</td>
<td>201</td>
</tr>
<tr>
<td>G.6 Zeta</td>
<td>205</td>
</tr>
<tr>
<td>G.7 Eta</td>
<td>209</td>
</tr>
<tr>
<td>G.8 Theta</td>
<td>213</td>
</tr>
<tr>
<td>G.9 Iota</td>
<td>217</td>
</tr>
<tr>
<td>H Lecturer interview case studies</td>
<td>220</td>
</tr>
<tr>
<td>H.1 Kappa</td>
<td>220</td>
</tr>
<tr>
<td>H.2 Lambda</td>
<td>227</td>
</tr>
<tr>
<td>H.3 Mu</td>
<td>232</td>
</tr>
<tr>
<td>H.4 Nu</td>
<td>241</td>
</tr>
<tr>
<td>H.5 Xi</td>
<td>250</td>
</tr>
<tr>
<td>H.6 Omicron</td>
<td>257</td>
</tr>
<tr>
<td>References</td>
<td>263</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>An illustration of the process of formative assessment with respect to a task (Yorke, 2003, p. 487)</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>The hierarchy of activity</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>The structure of an activity from Engeström (1999)</td>
<td>48</td>
</tr>
<tr>
<td>3.3</td>
<td>An elementary learning cycle</td>
<td>58</td>
</tr>
<tr>
<td>3.4</td>
<td>An Engeström representation of a student undergoing a CAA test</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Activity in Marx (1999) as parts of the Engeström (1987a) triangle</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Adding assessment and feedback to the learning cycle</td>
<td>70</td>
</tr>
<tr>
<td>3.7</td>
<td>The model for effective assessment</td>
<td>71</td>
</tr>
<tr>
<td>5.1</td>
<td>Alpha was largely satisfied that the CAA tests helped her to get the outcome she aimed for.</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>However, she also felt that the system did not always help to identify her weaknesses.</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>Learning cycles for Alpha</td>
<td>97</td>
</tr>
<tr>
<td>5.3</td>
<td>Epsilon was motivated towards marks and understanding the material in his motive to become a</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>pilot. No obvious contradictions arose in his interviews.</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>Kappa’s activity for assessment showing contradictions sited within the subject and between</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>the tools and object of activity</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>The assessment is effective for Kappa’s goal to assess students quickly on their procedural</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>knowledge, but it is not effective for testing conceptual knowledge</td>
<td></td>
</tr>
<tr>
<td>A.1</td>
<td>An example of one of the questions in the CAA system</td>
<td>158</td>
</tr>
<tr>
<td>A.2</td>
<td>An example of a results page in the CAA system</td>
<td>159</td>
</tr>
<tr>
<td>A.3</td>
<td>An example of the feedback given by the CAA system</td>
<td>159</td>
</tr>
<tr>
<td>F.1</td>
<td>Students’ perceptions of the importance of mathematics in their future studies, separated by</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>cohort</td>
<td></td>
</tr>
<tr>
<td>F.2</td>
<td>Students’ perceptions of the importance of mathematics in their future careers, separated by</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>cohort</td>
<td></td>
</tr>
<tr>
<td>F.3</td>
<td>The amount of mathematics that students are willing to engage in on their courses, separated</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>by cohort</td>
<td></td>
</tr>
<tr>
<td>F.4</td>
<td>The number of times students normally access the practice tests, separated by cohort.</td>
<td>176</td>
</tr>
<tr>
<td>F.5</td>
<td>Students’ selections of words they associated with computer-aided assessment, both before</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>experiencing CAA and at the end of the year</td>
<td></td>
</tr>
<tr>
<td>F.6</td>
<td>The extent to which the students felt that it was easy to achieve 100% in the summative test</td>
<td>178</td>
</tr>
</tbody>
</table>
F.7 The degree classifications and summative test scores that the students aimed for. .................................................. 179
F.8 The students' affinity towards group work. ................................. 179
F.9 Students' responses when asked the extent to which they agree three aspects of their knowledge was being tested by the summative test. .... 180
F.10 A comparison between what students believe an assessment should achieve and what students believe CAA does. .......................... 181

G.1 Alpha was largely satisfied that the CAA tests helped her to get the outcome she aimed for. However, she also felt that the system did not always help to identify her weaknesses. ................................. 183
G.3 Beta’s motives changed between semesters. It had a small effect on her activity as a whole, as she tried harder to achieve 100% in her CAA tests in the second semester. .............................................. 189
G.4 Beta’s learning cycles for the first semester. ............................... 191
G.5 Beta’s learning cycles for the second semester. ............................ 192
G.6 Gamma desired to achieve the best possible degree classification and believed that achieving 100% in the CAA summative tests was realistic. His views on collaboration and cheating were different to many of his peers. 194
G.8 Delta’s collaborations with peers was contrary to her views on collaboration more generally. She found marks to be rewarding and thus her actions were oriented towards the pursuit of marks. .............................. 198
G.9 Delta’s learning cycles. .......................................................... 199
G.10 Epsilon was motivated towards marks and understanding the material in his motive to become a pilot. No obvious contradictions arose in his interviews. ................................................................. 202
G.11 Epsilon’s learning cycles. ....................................................... 204
G.12 Zeta used the CAA tests to evaluate his current understanding; however he felt that getting zero marks for an incorrect answer did not help him. 206
G.13 Zeta’s learning cycles. .......................................................... 207
G.14 Eta appeared to get more out of the CAA tests than most; however he found the system could not help him beyond his initial learning goal to get 100% in the summative test. ................................. 209
G.15 Eta’s learning cycles. ............................................................ 211
G.16 Theta relied much less on her student peers than other students in the study. She encountered a contradiction when the lecturer changed after the end of the first semester. ............................................ 213
G.17 Theta’s learning cycles for the first semester. ............................. 215
G.18 Theta’s learning cycles for the second semester. .......................... 216

H.1 Kappa’s activity for assessment showing contradictions sited within the subject and between the tools and object of activity .................. 220
H.2 The assessment is effective for Kappa’s goal to assess students quickly on their procedural knowledge, but it is not effective for testing conceptual knowledge ......................................................... 226
H.3 Lambda’s activity involves the use of paper tests as well as the CAA practice tests to assess his students. He placed additional rules for assessment on himself, taking into consideration his experiences and beliefs of best practice. ................................. 226
H.4 Lambda found the combination of practice tests and paper tests to be effective, though it is not clear how this practice will develop . . . . . . . 231

H.5 Mu used CAA as a means to encourage her students to practise between lectures. The pitfalls in using CAA as a summative assessment meant that she did not regard students’ scores as important. . . . . . . . 232

H.6 Mu found the CAA effective enough for her main goals: to manage large-scale assessment and getting students to practise using procedures. It was not effective for assessing conceptual skills. . . . . . . . . . . . . 240

H.7 Nu used CAA within an assessment regime in which he sought to change students’ interpretations of what it means to “do mathematics”. The CAA system proved to be ineffective for this aim. . . . . . . . . . . . . 241

H.8 Mu found the CAA effective enough for her main goals: to manage large-scale assessment and getting students to practise using procedures. It was not effective for assessing conceptual skills. . . . . . . . . . . . . 249

H.9 Xi wished to use CAA to offer students timely and encouraging feedback; however, she found that the tool did not adequately assist her in this task. 250

H.10 Xi’s main aims were to provide timely yet encouraging feedback to students so that they gain confidence and progress. She found that it did not do so for weaker students. . . . . . . . . . . . . . . . . . . . . . 255

H.11 Omicron had two aims for assessment: to encourage students to practise and to give marks to students. The latter aim was imposed by the community rules. She felt that the CAA system questions were not sufficient for her needs. . . . . . . . . . . . . . . . . . . . . . 256

H.12 CAA was effective for Omicron’s two main goals for assessment; however she felt the questions were not challenging enough to get a good appreciation of her students’ conceptual knowledge. . . . . . . . . . . . . . 262
List of Tables

2.1 Aspects of formative assessment (Black & Wiliam, 2009, p. 8) .......... 18
2.2 Comparison of aspects of assessment and feedback from three perspectives 20
3.1 Examples of prerequisites that invoke goal formation in CAA .......... 55
3.2 Further examples of prerequisites of goal formation in CAA .......... 56
3.3 Levels of contradictions and corresponding learning actions, from Foot (2014, p. 342) ......................................................... 67
4.1 Student interview themes identified from the questionnaires .......... 85
5.1 A breakdown of the students that completed both questionnaires and those that took part in interviews ................................... 93
5.2 Students’ relationships with mathematics — ✓ indicates confirmation during the interview, ✗ indicates denial, blank indicates no references made during interview ................................................. 99
5.3 Students’ practice test aims ................................................. 104
5.4 Practice test actions ....................................................... 107
5.5 Use of collaboration and peer support .................................. 111
5.6 Causes of the end of the learning cycle ................................ 116
5.7 Lecturer practices for the use of CAA ................................ 120
5.8 Lecturers’ reasons for using CAA ..................................... 121
5.9 Lecturers’ aims for assessment ......................................... 131
5.10 Contradictions that lecturers faced when using CAA ................. 137
5.11 Lecturers’ interpretations of collaboration in CAA testing .......... 142
5.12 Lecturers’ aims that were satisfied and not satisfied by the CAA system 144
F.1 The distribution of questionnaire responses across cohorts .......... 172
Abbreviations

AIM  Alice Interactive Mathematics
CAA  Computer-Aided Assessment
CAS  Computer Algebra System
CBA  Computer-Based Assessment
CHAT Cultural-Historical Activity Theory
FA  Formative Assessment
FI  Feedback Intervention
HEI  Higher Education Institution
HELM Helping Engineers Learn Mathematics
HESA  Higher Education Statistics Agency
ICT  Information and Communication Technology
IT  Information Technology
JISC  Joint Information Systems Committee
MCQ  Multiple Choice Question
OMR  Optical Mark Recognition
RQ  Research Question
STEM Science, Technology, Engineering and Mathematics
Dedicated to Mum
Chapter 1

Introduction

A computer-aided assessment (CAA) system had been used at a higher education institution for more than ten years, but had not been formally evaluated. It was used with first year students that were studying mathematics modules towards mathematics or engineering degrees. The question bank for engineering students was the product of the HELM (Helping Engineers Learn Mathematics) Project (Green, Harrison, Palipana, Pidcock, & Ward, 2004); a second question bank was later developed for the same system, but for mathematics students.

The need for an evaluation arose from anecdotal concerns that students were not using the CAA tests in the manner in which it was intended. In particular, some lecturers were suspicious that some students were cheating — for example, some students had managed to complete the test in less than one second — or they were using the model feedback on offer in the practice tests to model their solutions in the summative test.

An earlier focus group showed that students were particularly dedicated to the CAA tests in order to maximise their marks; they would use the practice tests to perfect their methods ready for the summative test (Broughton, Hernandez-Martinez, & Robinson, 2012). However, those students also felt that some students were able to do well in the tests without engaging with the course (Broughton, Hernandez-Martinez, & Robinson, 2011).

These concerns made an evaluation of the CAA system necessary; however, evaluations of assessment tools appear not to be common. Indeed, the literature review (Chapter 2, p. 4) suggests that there is no rigorous framework from which to evaluate an assessment, though the literature does suggest qualities and characteristics of a desirable assessment (for example, Gibbs & Simpson, 2005; Nicol & Macfarlane-Dick, 2007; Black & Wiliam, 2009).
Chapter 1  Introduction

The second part of the literature review discusses the use of CAA to assess mathematics in higher education (p. 27). Its use has increased substantially since primitive forms of CAA were introduced in the 1970s and are now relatively commonplace in higher education mathematics courses. There are diverse systems in current use and continue to be developed. Some contemporary systems use underlying mathematical software to judge algebraic and graphic responses to questions, while some systems were reliant on multiple choice or numerical input questions — such as the system evaluated for this study.

The literature review concludes that the need for an evaluation of a CAA system is not unique to the situation in this study: there is a lack of a theoretically-conceived evaluation framework for assessing and critiquing assessment tools — Taras (2010a) notes this with regard to assessment for learning and formative assessment; Tight (2004) comments on the lack of theorisation in higher education research more widely. The issues that were identified anecdotally seemed to be complaints made more generally in the higher education sector and there is a need for a framework to assess assessment.

The need for an evaluation, and with a need for a framework for evaluating the CAA system, led to the three research questions (RQs) for this study.

**RQ1.** What makes an assessment “effective”?

**RQ2.** Under what conditions is the CAA system an effective assessment?

**RQ3.** What would make the CAA system more effective as an assessment tool?

To build a framework for assessment, the theoretical framework chapter (Chapter 3, p. 45) provides a basis from which to examine activity and helps identify what an assessment would ideally achieve. Cultural-historical activity theory (CHAT) is a framework that models human activity and provides points of reference for analysing how humans act. CHAT is particularly useful in relating actions to their intentions — in CHAT terminology, goals direct actions. Goals are important in assessment; they are what students and lecturers wish to achieve. This motivates the model for effective assessment (p. 68). This model provides a framework to determine the scope of effectiveness for an assessment.

The methodology chapter (Chapter 4, p. 73) discusses how to obtain data to determine where the assessment was effective and where it ceased to be effective, with respect to both students and lecturers. The theoretical framework directs the methodology to an extent, since it is necessary to obtain information about an individual’s goals and actions.
This chapter explains the reasons for using questionnaires and interviews to obtain these data, outlines how these methods were used, and how the data were analysed.

The questionnaires and interviews with students and lecturers provided individual case studies for students and lecturers using the CAA system for assessment. From their perspectives, the analysis chapter (Chapter 5, p. 91) determines the activity of assessment through this CAA system. The model for effective assessment is used to show that the CAA system was effective for some aspects of what students and lecturers wanted to achieve through an assessment exercise, but there were substantial limitations that prevented the system from being more effective.

This thesis concludes with some reflections on using the model for effective assessment and how the individuality of the users of CAA mean that the assessment can never be universally and always effective. The conclusions chapter (Chapter 6, p. 153) makes some recommendations for the use of CAA system, based on the practices that the students and lecturers reported in the interviews, that appear to increase its scope of effectiveness.
Chapter 2

Literature Review

The primary aim of this project was to evaluate a mathematics-based computer-aided assessment (CAA) system that has been used for several years with first year students at a higher education establishment. This literature review starts by making several points:

- There is little guidance in the literature for performing evaluations on an assessment.
- Existing frameworks and definitions are not universally accepted in the literature.
- There is an opportunity to develop a model for an effective assessment using existing literature and a theoretical framework.

This literature review examines higher education assessment literature and the use of CAA in mathematical sciences in higher education. The methods for searching for articles relating to these two aspects of assessment included the use of keywords in Google Scholar and Loughborough University’s library website search facility. These keywords included:

- assessment for learning
- CAA
- computer-aided assessment
- computer-based assessment
- engineering mathematics
- feedback
- formative assessment
- higher education
- summative assessment

From the papers identified using combinations of these search strings, those referring to the use of CAA systems in higher education mathematics and engineering mathematics — or seminal papers on higher education assessment — were considered in this literature review.

One of the aims of this study, and this literature review, is to address the first research question (RQ) in order to determine and justify whether the CAA system is effective.
RQ1. What makes an assessment tool effective?

The second part of this literature review provides a background to CAA. While the CAA system cannot be evaluated using existing literature, it is important for the reader to understand why CAA is becoming far more relevant in mathematics education at higher education level and its implications on teaching and learning practice.

The literature reports of increasing uptake of CAA in mathematics-based subjects (Greenhow, 2015, for example); and that doing so brings many benefits, both for lecturers and for students (Bull & Danson, 2004). However, there also issues with using CAA. This literature review will discuss both the benefits and the problems.

There have been notable advances in mathematics-based CAA systems since the turn of the century (Sangwin, 2013); and as such systems become more readily available, the evaluation of the existing system at the studied institution is ever more pertinent. This literature review examines how the current system compares against more contemporary systems, seeking to address the second and third research questions:

RQ2. Under what conditions is the CAA system an effective assessment?

RQ3. What would make the CAA system more effective as an assessment tool?

2.1 Making assessment effective

Making assessment effective is a challenge — not least since it is not immediately clear what it means for an assessment to be effective. Nonetheless, the literature provides some examples where assessment has been effective and what their effects are. More theoretical literature provides definitions of and criteria for good assessment practice. Formative assessment is a prominent term in the literature and its proponents argue that, when employed correctly, is effective. This section of the literature review examines assessment practices and their effects, as applied to any level of education, but interpreted with regard to higher education and, where appropriate, to the assessment of mathematics at that level.

Though the aim of this first part of the literature review is to establish what makes assessment effective, it is useful to examine current assessment practices in mathematics courses in higher education in the United Kingdom, since it is in this situation the existing CAA system resides.
2.1.1 Current assessment practice

Closed book examination is, by some distance, the dominant method for mathematics assessment in higher education (Iannone & Simpson, 2012): it contributes over three quarters of the final mark in nearly 70% of university mathematics modules. In their discussion of this study, they concluded:

“The image developed by the survey is of a system dominated by the closed book examination, but — from the point of view of heads of department — not inappropriately so. The closed book examination is seen to have an appropriate balance of validity, efficiency, reliability and fairness.” (Iannone & Simpson, 2012, p. 13)

The lack of desire for change to new assessment practices is documented elsewhere in the literature. They seem to confirm the findings of Iannone and Simpson (2012) in so much that closed book examinations are well-established and are highly regarded. Sunal et al. (2001), drawing on previous research, listed several causes for a lack of change of teaching in higher education and this is not a phenomenon specific to mathematics. At the institutional level, they suggested that the culture, the lack of professional development, and the organisational context and structure lead to ineffective teaching and the suppression of change. For individual instructors, Sunal et al. added, citing Cuban (1990), that lecturers’ prior beliefs, expectations and perceptions of the limitations of their teaching contexts prevented change to teaching practice. They added:

“The most commonly ranked barriers to course level change accounted for 60% of those identified very important by faculty members. They were resources, time, and turf conflicts. Faculty members perceived themselves as having little control over these areas.” (Sunal et al., 2001)

With respect to mathematics assessment, Smith and Wood (2000, p. 131) declared that “there is an innate conservatism of many mathematicians”, for whom a lack of personal experience using new assessment practices invokes a resistance to using them. They also suggested that many students are averse to tasks that require knowledge beyond routine and procedure.

Institutionally, Smith and Wood added that time and financial constraints also hamper change. Taras (2002, p. 501) noted “that innovation in assessment [is] no longer an option in higher education” because “staff are deterred from attempting to implement innovations and ideas which do not conform to the protocols of internal and external quality
control”. A common theme in the literature is the seemingly contemporary problem of managing limited resources. Both Sunal et al. (2001) and Smith and Wood (2000) alluded to this issue; Gibbs and Simpson (2005, p. 9) reported that such constraints have already “led to a reduction in the frequency of assignments, in the quantity and quality of feedback and in the timeliness of this feedback.” Baker (1999) mentioned cost as one of several demands on assessment, which ultimately inhibits change.

Therefore, there are a number of factors that prevent assessment practice from developing in line with the literature. The constraints placed upon innovation appear to be institutional and out of the control of lecturers. This has an impact on the effectiveness of an assessment, since lecturers may have aims that relate to developing assessment in accordance with emerging best practices in the literature.

The literature provides many reasons for developing assessment practice in higher education. Taras (2002) argued that current assessment practices undermine students’ autonomy and independence, contrary to the warning given by the Dearing Report (Dearing, 1997) that individuals entering work need to develop such qualities.

Yorke (1998) suggested there were many areas for improvement in assessment practice at the time, particularly noting the lack of consistency in practice and the tensions that exist between institutional desires for standardisation and the discretion of departments to employ their own methods.

Gibbs and Simpson (2005) refuted the claim that coursework is a less valid assessment than examinations:

“Higher average marks and student preference would not count for much if coursework were inherently less valid as an assessment — but it is not. First, examinations are poor predictors of any subsequent performance, such as success at work. . . .

“Second, coursework marks are a better predictor of long term learning of course content than are exams. . . .

“Third, in experimental studies in which students have either studied exam-based or assignment-based courses, the quality of their learning has been showing to be higher in the assignment-based courses.” (Gibbs & Simpson, 2005, p. 7)

They add that “assessment sometimes appears to be, at one and the same time, enormously expensive, disliked by both students and teachers, and largely ineffective in supporting learning” (Gibbs & Simpson, 2005, p. 11). Worse still, “it is a common observation of higher education teachers that if coursework is taken away from a module due
to resource constraints, students simply do not do the associated studying” (Gibbs & Simpson, 2005, p. 8).

This section of the literature found that there are many barriers to developing and changing assessments. There are many constraints on lecturers in higher education that restrict the extent to which they can develop their assessment practice; many of these constraints were imposed by the department or the university, which control the budget, the time available, the resources available and the teaching spaces. There are opportunities to develop assessment, but it is unclear from the literature whether these constraints are insurmountable, or whether there are other reasons why lecturers do not examine other methods of assessment. The fact that mathematics assessment at undergraduate level is largely dependent on examinations (Iannone & Simpson, 2011), demonstrates the lack of development and diversity in undergraduate assessment in mathematics. The following section explores the various reasons for offering ongoing assessment during a course.

### 2.1.2 The role of the lecturer

The lecturer has a defining role in choosing, designing, delivering and marking assessments on the courses that they run. They choose and adapt assessment tools according to their needs; they select when and how the assessment tools should be accessed; and they are largely responsible for constructing feedback to be returned to the student. These make the lecturer an important agent in the effectiveness of an assessment.

This subsection explores why lecturers’ aims are important in the way that assessment is used, and how feedback potentially has an impact upon students’ capacity to find the assessment tool effective for their needs.

### The purpose of assessment

The reasons for performing an assessment differ greatly between lecturers. Yorke (1998, p. 108) believed there are three main purposes: “assisting in the process of learning; determining what learning has occurred; and providing evidence regarding the success or otherwise of the programme in question”. An assessment may be undertaken for one or more of these reasons; however, Yorke noted that much focus has been directed towards “determining what learning has occurred” at the expense of “assisting in the process of learning”.

Samuelowicz and Bain demonstrated the existence of such stark differences between the intentions of lecturers for the use of assessment in a variety of biology, medical and psychology courses. In their study, one academic viewed assessment “as a way of making
students study” (Samuelowicz & Bain, 2002, p. 186); another academic used assessment to encourage students to work harder and to differentiate between those capable of a professional standard; and a further lecturer believed that assessment is a tool for guiding students through learning and to help them achieve the professional standard.

Although the latter case appears to offer the most pedagogically supportive reasons for assessing students, the findings of Gibbs and Simpson (2005) that students tended to study less in the absence of coursework ought not to be neglected. This is particularly important in learning mathematics, as noted by Anthony and Knight (1999). In their discussion of existing studies of classroom practice and curriculum reform, they expressed concern that the elevated importance of understanding undermined the role of practice and repetition. They argued that practising should be considered distinct from rote learning and “drill practice”:

“In classrooms where learning is viewed [as] problem solving, where tasks are treated as problems to be solved rather than exercises to be completed using pre-specified procedures, practice should be viewed as a critical component of the development and maintenance of procedural knowledge and the understanding of mathematics.” (Anthony & Knight, 1999, p. 103)

One might argue that the role of practice in mathematics learning is higher than in other subjects. This claim is supported by a number of papers and books that have lauded the benefits of analysing students’ mistakes — to guide learning and improvement by identifying common errors and misconceptions (Ryan & Williams, 2007; Radatz, 1979; “Pedagogic practices and interweaving narratives in AS Mathematics classrooms”, 2007, for example). In this way, students should have the opportunity to learn and develop in response to feedback on the learning they have already encountered.

This section described many uses for assessment in the process of learning. Assessment provides opportunity for students to practise using the content they have learned in lectures and gain feedback on this work. Assessment also provides feedback to the lecturer, who can gain an impression of the level of understanding among the student cohort. It also prepares students for their future profession.

It is not clear from the literature to what extent lecturers consider all these aims when using assessment with these students. Consequently, it does not indicate how lecturers’ aims influence which assessment techniques they use and how they use them. This is an important consideration, since an assessment is effective to the lecturer if it accomplishes the aims of the lecturer. Therefore, the aims of the lecturer are important to this study.
Searching for best practice in offering feedback

Feedback, with regard to assessment, is information produced as a result of assessment that documents a person’s performance on a task, which forms the basis for improvement. Feedback is necessary for some lecturers’ aims, potentially — particularly with regard to encouraging students to develop their skills. Likewise, it may also be important for students’ aims for assessment.

Undoubtedly, feedback can have a positive contribution towards students’ learning; however, it is by no means guaranteed that providing feedback will lead to an improvement for many reasons. Nicol and Macfarlane-Dick (2007) expressed the problem in the literature as a failure in the past to advise practitioners of best practice:

“The literature on external feedback is undeveloped in terms of how teachers should frame feedback comments, what kind of discourse should be used, how many comments are appropriate and in what context they should be made. Much more research is required in this area.” (Nicol & Macfarlane-Dick, 2007, p. 209)

Consequently, as the papers by Nicol and Macfarlane-Dick (2007) and by Kluger and DeNisi (1996) argued, feedback can have negative effect on progress, motivation and self-esteem on students — as well as a positive one. Hattie and Timperley (2007) believed that students are too often given feedback with the intention to motivate the student, but lacks direction or advice; or as Hattie and Timperley (2007, p. 88) stated, the feedback fails to answer the student’s questions, “where am I going?”, “how am I going?”, and “where to next?”.

Kluger and DeNisi (1996), in their review of feedback practice literature spanning one hundred years, reported that, for much of that time, the literature gave a misleading impression of the effects of feedback:

“We argue that a considerable body of evidence suggesting that feedback intervention (FI) effects on performance are quite variable has been historically disregarded by most FI researchers. This disregard has led to a widely shared assumption that FIs consistently improve performance.” (Kluger & DeNisi, 1996, p. 254)

In response to Kluger and DeNisi, Wiliam (2006) highlighted the extent of the problem that was hidden in the literature: that in 40% of the studies, students’ performances
were lower with feedback than they would have achieved without. Wiliam described the nature of the problem:

“The key feature of these studies was that . . . the feedback focused attention on the person rather than the quality of the work. . . . For the 60% of studies that found a positive impact on performance, Kluger and DeNisi found that the biggest impacts occurred when feedback told not just what to do to improve, but also how to go about it.” (Wiliam, 2006)

Hattie and Timperley (2007) offered “a model of feedback to enhance learning”, which included four levels at which feedback is aimed:

- **Task level** — How well tasks are understood/ performed;
- **Process level** — The main process needed to understand/perform tasks;
- **Self-regulation level** — Self-monitoring, directing, and regulating of actions;
- **Self level** — Personal evaluations and affect (usually positive) about the learner.

Hattie and Timperley (2007, p. 87)

They noted that “feedback at the self or personal (usually praise) . . . is rarely effective” (Hattie & Timperley, 2007, p. 102): an echo of Wiliam’s 2006 conclusions from the findings of Kluger and DeNisi. Feedback, then, ought to be aimed towards developing students’ capabilities in terms of tasks, processes and self-regulation.

Addressing feedback towards the task level is, as Hattie and Timperley (2007, p. 93) described, related to “surface learning” — that is, being able to carry out procedures — whereas feedback about the processing of the task is connected to deeper learning — for example, making connections between topics, applying knowledge to new situations. They added: “feedback at the process level appears to be more effective than at the task level for enhancing deeper learning” (Hattie & Timperley, 2007, p. 93). What it means for feedback to be effective is discussed later.

Feedback on the self-regulation level guides the student towards independence and becoming more “effective learners” (Hattie & Timperley, 2007, p. 94). In doing so, students rely less upon their lecturers or the tools of the task to obtain feedback and develop their own strategies for improving performance. An example of such self-regulation is the ability to identify errors in solutions independently.
2.1.3 The role of the student

It is desirable for students to develop self-regulation and independence. Hattie and Timperley (2007, p. 101) noted that “students, too often, view feedback as the responsibility of someone else, usually teachers, whose job it is to provide feedback information by deciding for the student how well they are going, what the goals are, and what to do next”.

It is not only students that believe that the success of an assessment is the responsibility of the lecturer. Wiliam (2006) argued that teachers are under pressure to induce learning in order to improve results. He added: “Only learners create learning, and so, when we look at the role that assessment plays in promoting learning, the crucial feature is not the validity of the assessment, or its reliability, but its impact on the student” (Wiliam, 2006, p. 4).

For some, the student must have an active role within the definition of feedback. Ramaprasad (1983), cited in Sadler (1989, p. 120), defined feedback in the following way: “Feedback is information about the gap between the actual level and reference level of a system parameter which is used to alter the gap in some way”. Sadler’s interpretation of this definition is that “information about the gap between actual and reference levels is considered as feedback only when it is used to alter the gap” (Sadler, 1989, p. 120, emphasis in original).

Requiring that the student responds to feedback seems a certain requirement if the feedback is to be considered effective. Gibbs and Simpson (2005, p. 8) reported that “feedback is often not read at all…or not understood”. They cite an article by Wotjas (1998), in which some students do not acknowledge the importance of feedback at all:

“While many students improve their work when they understand the purpose of feedback and know more about the assessment criteria, other (sic) still see it as an unhelpful burden.

...”

“Some students threw away the feedback if they disliked the grade, while others seemed concerned only with the final result and did not collect the marked work.” (Wotjas, 1998)

Placing the responsibility upon and requiring action from the student, Sadler (1989, p. 121) set three conditions: the student has to “(a) possess a concept of the standard…being aimed for, (b) compare the actual…level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap”, forming three “necessary conditions, which must be satisfied simultaneously”.

Although Struyven, Dochy, and Janssens (2005) were not focussing on mathematics assessment in particular, they found that students were not often satisfying these three conditions; when lecturers were seeking deeper learning in their students, many students persisted with surface learning. Furthermore, they noted that students tended to approach assessment by seeking to fulfil the explicit requirements of the task, rather than go beyond them.

“They do not look for the meaning of a text or an assignment, understanding is not their purpose. They focus on details, try to memorise parts, and study the layout, with the purpose of meeting the requirements of the task (or evaluation). Some students appear not to gain insight into texts. These findings prompt the following questions: do they prepare their examinations in the same way? As educators, do we have (any) influence on their studying behaviour and approaches to learning?” (Struyven et al., 2005, p. 334)

They further argued that students approach assessment in the way that they conceive learning and, conversely, the results of assessment direct their future learning. Consequently, “Assessment is thus logically, but also empirically, one of the defining features of students’ approaches to learning” (Struyven et al., 2005, p. 332). For some at least, students take cues from assessment and, therefore, assessment has an important role in students’ learning.

2.1.4 A definition for effective assessment

In seeking a definition for “effective assessment”, the roles and qualities of both the lecturer and the student have been explored; the level at which feedback is offered has been addressed; and the purpose of the assessment has been discussed. It would appear that these would form the basis for an assessment to satisfy the needs of both the students and the lecturer. These aspects of an assessment are important when considering whether an assessment could be effective or not.

However, any definition for effective assessment is likely to be flawed. As Gibbs and Simpson (2005, p. 11) warned, “The evidence is rarely conclusive enough to argue that if your assessment fulfils these conditions then learning will inevitably be more effective.”

Nonetheless, there are guidelines in the literature for feedback or assessment that aim to make assessment effective. Hattie and Timperley (2007, p. 86) described effective feedback as feedback that answers the questions asked by students or teachers: “Where
am I going? (What are the goals?), How am I going? (What progress is being made towards the goal?), and Where to next? (What activities need to be undertaken to make better progress?).

Thus, Hattie and Timperley believe that students must be at the centre of feedback; there must be a shared awareness of assessment goals; and there has to be an indication of current performance and what is needed to improve this performance. However, as Struyven et al. (2005) noted, many students find the “Where to next?” question a challenge and are guided by the assessments they are given.

Although Nicol and Macfarlane-Dick did not describe their seven principles of good feedback practice as guidelines for effective feedback, their principles followed similar themes. They believed that good feedback practice:

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality teaching information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape teaching.

(Nicol & Macfarlane-Dick, 2007, p. 205)

Beyond the themes present in the definition of effective feedback from Hattie and Timperley (2007), these seven principles offer some additional aspects to the assessment and feedback processes. Point 2 has implications for lecturers: assessments would have to be designed to include opportunities for students to self-assess and reflect upon their learning. Point 4 encourages lecturers and students to “conceptualise feedback more as dialogue rather than as information transmission” (Nicol & Macfarlane-Dick, 2007, p. 210, emphasis in original), in more of a dialogic approach (see Bakhtin, 1981) rather than in the transmissionist tradition.

Points 6 and 7 serve to act as evidence that feedback has been effective: that is, the student should be able to demonstrate his or her improvement in performance; and the lecturer should be aware of weak areas in his or her cohort that need to be addressed in further instruction.
Gibbs and Simpson (2005) offer ten conditions of assessment and feedback that affect students’ engagement and quality of learning. They are:

**Influences of assessment on the volume, focus and quality of studying**

1. Sufficient assessed tasks are provided for students to capture sufficient study time.
2. These tasks are engaged with by students, orienting them to allocate appropriate amounts of time and effort to the most important aspects of the course.
3. Tackling the assessed task engages students in productive learning activity of an appropriate kind.

**The influence of feedback on learning**

4. Sufficient feedback is provided, both often enough and in enough detail.
5. The feedback focuses on students’ performance, on their learning and on actions under the students’ control, rather than on the students themselves and on their characteristics.
6. The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance.
7. Feedback is appropriate to the purpose of the assignment and to its criteria for success.
8. Feedback is appropriate, in relation to students’ understanding of what they are supposed to be doing.
9. Feedback is received and attended to.
10. Feedback is acted upon by the student.

(Gibbs & Simpson, 2005, p. 12–24)

In these conditions, six points relate to the actions of the lecturer — he or she must design tasks that demand sufficient study time; and provide opportunities for frequent, timely and detailed feedback. As for students, they must spend sufficient time on such assignments — and use this time appropriately — and respond to the feedback they receive.
Relating conditions for effective assessment with formative assessment

The themes and conditions raised thus far provide a useful basis upon which to explore formative assessment. The premise of formative assessment is to provide feedback to students at each stage of the learning process, where summative assessment occurs at the end of the learning process (Bennett, 2011, p. 6). However, quite how formative assessment is defined, its theoretical basis, and whether a particular assessment technique can claim to be “formative assessment” is cause for much debate in the literature. Of particular interest, however, is whether formative assessment is an effective assessment.

Formative assessment is pertinent to the exploration of effective assessment since formative assessment is described as effective in Black and Wiliam (1998). They conducted an extensive review of existing literature pertaining to use of formative assessment at the time. Though much of the paper focused on secondary education level, they included some studies conducted at undergraduate level.

It does not necessarily follow that formative assessment is “effective assessment”, however. Yorke (1998, p. 106) ensured that one definition of effectiveness is not forgotten: “a well-constructed system for the management of assessment will ensure that what is expected to take place actually does take place (i.e. it is effective)” (Yorke, 1998, p. 106, emphasis added). That is, effect sizes demonstrate an improvement on the part of student learning, but the assessment must still satisfy the aims of the student and the lecturer.

Although the notion of a difference between assessments with formative or summative intentions has a longer history than implied here, Sadler (1989) and Ramaprasad (1983) are most credited with advancing progress towards a definition for formative assessment. Ramaprasad (1983, p. 4) defined formative feedback as “information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way”. Sadler used Ramaprasad’s definition to describe the process of formative assessment. Sadler stated:

“Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student’s competence by short-circuiting the randomness and inefficiency of trial-and-error learning.” (Sadler, 1989, p. 120)

The author believes that by “short-circuiting the randomness of inefficiency of trial-and-error learning”, Sadler meant that formative assessment seeks to provide guidance to students that are not yet able to reach a solution in the most appropriate or efficient manner.
Sadler, too, believed that students have an active role in the assessment process, noting that Ramaprasad believed that “information about the gap between actual and reference levels is considered feedback only when it is used to alter the gap” (Sadler, 1989, p. 121, emphasis in original). Furthermore:

“...the learner has to (a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap.” (Sadler, 1989, p. 121, emphases in original)

Finding a definition of formative assessment in the literature is not so straightforward, despite the growing body of literature devoted to it (Taras, 2005, p. 466). Often, only aspects of formative assessment are alluded to: in Black and Wiliam (1998), assessment was defined as an activity that provides feedback to adapt teaching and learning; it becomes formative assessment when “the evidence is actually used to adapt the teaching work to meet the needs” (Black & Wiliam, 1998, p. 2). This definition lacks the implication of student activity. Taras (2010b, p. 127) offers a simple definition of formative assessment — “a summative assessment produces feedback which when used results in formative assessment” — which lacks both the nature of the feedback and the nature of the “results”.

What is different about the definition offered by Taras (2010b) is that formative assessment requires the presence of a summative assessment, which is presented as a “judgement... at that point in time” (Taras, 2010b, p. 127). This seems to contradict Bennett’s (2011) interpretation, in which summative assessment occurs at the end of the learning process. As Taras (2005, p. 466) noted, “assessment is of central importance in education, and yet there is a lack of commonality in the definition of terminology relating to it” — it is evident from such contrasting views that this lack of commonality remains a problem.

Formative assessment does not have a definition agreed by consensus in the literature. It is characterised as a group of processes and conditions: the common aspect in all of these characterisations is the presence of feedback aimed at improving the student. Other aspects of formative assessment are disputed: in particular, the role of the student in deciding learning goals and whether the student must improve as a consequence of receiving feedback are not present in all descriptions of the formative assessment process.

The “aspects of formative assessment”, as described by Black and Wiliam (2009, p. 8) in Table 2.1, the requirements of teachers, peers and learners. The columns of this
Table 2.1: Aspects of formative assessment (Black & Wiliam, 2009, p. 8)

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Where the learner is going</th>
<th>Where the learner is right now</th>
<th>How to get there</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clarifying learning intentions and criteria for success</td>
<td>2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding</td>
<td>3. Providing feedback that moves learners forward</td>
<td></td>
</tr>
<tr>
<td>Peer</td>
<td>Understanding and sharing learning intentions and criteria for success</td>
<td>4. Activating students as instructional resources for one another</td>
<td></td>
</tr>
<tr>
<td>Learner</td>
<td>Understanding learning intentions and criteria for success</td>
<td>5. Activating students as the owners of their own learning</td>
<td></td>
</tr>
</tbody>
</table>

table match the three questions that effective feedback answers according to Hattie and Timperley (2007, p. 86) (though this is not explicitly stated by Black and Wiliam).

The assumption, then, is that formative assessment should reflect these five aspects:

1. Clarifying and sharing learning intentions and criteria for success;
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
3. Providing feedback that moves learners forwards;
4. Activating students as instructional resources for one another; and
5. Activating students as the owners of their own learning.

(Black & Wiliam, 2009, p. 8)

Nicol and Macfarlane-Dick (2007, p. 205) referred to peer dialogue around learning in their seven principles of good feedback practice. Black and Wiliam appeared to go further: peers have a role in sharing goals and mutual assessment. Unfortunately, Black and Wiliam did not elaborate much further on peer assessment: only that using learners’ feedback to motivate further learning is a complex process.

There are some links between the five criteria for formative assessment proposed by Black and Wiliam (1998), the seven criteria for good feedback practice in Nicol and Macfarlane-Dick (2007), and the ten conditions under which students learning is supported by assessment by Gibbs and Simpson (2005). These conditions could be synthesised (as in Table 2.2) to manifest in the self, assessment, feedback, student response and interaction levels.
Comparing these criteria in this way demonstrates how it is not only the design of the assessment tool that makes the process of assessment an ideal one. It develops the individual undergoing assessment (the self level), it helps to form learning goals (the goal level), it produces useful feedback from which the student can develop (the feedback level), it necessitates a reaction from the student to develop (the response level) and it fosters communication between students, peers and teachers (the interaction level) — as well as informing the design of the assessment tool itself (the assessment level).

Formative assessment seems to provide a simple impression of what it is to be effective assessment, given its demonstrable effect sizes (Black & Wiliam, 1998) and the fact it encompasses many of the themes expressed elsewhere — most notably in Table 2.2. However, there are some problems with using formative assessment as a standard for assessment.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Encourages positive motivational beliefs and self-esteem</td>
<td>Understanding learning intentions and criteria for success</td>
<td></td>
</tr>
<tr>
<td><strong>Goal level</strong></td>
<td>Help clarify what good performance is (goals, criteria, expected standards)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment level</strong></td>
<td>Sufficient assessed tasks are provided for students to capture sufficient study time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>These tasks are engaged with by students orienting them to allocate appropriate amounts of time and effort to the most important aspects of the course</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tackling the assessed task engages students in productive learning activity of an appropriate level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------</td>
<td>----------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Sufficient feedback is provided both often enough and in enough detail</td>
<td>Delivers high quality information to the student about their learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The feedback focuses on students’ performance, on their learning and on actions under the students’ control, rather than on the students themselves and on their characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The feedback is timely in that it is received by student while it still matters to them and in time for them to pay attention to further learning or receive further assistance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback is appropriate to the purpose of the assignment and to its criteria for success</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback is appropriate, in relation to students’ understanding of what they are supposed to be doing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Response level</strong></td>
<td>Feedback is received and attended to Feedback is acted upon by the student</td>
<td>Facilitates the development of self-assessment (reflection) in learning</td>
<td>Provides opportunities to close the gap between current and desired performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provides information to the teacher that can be used to help shape teaching</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction level</strong></td>
<td></td>
<td>Encourages teacher and peer dialogue around learning</td>
<td>Understanding and sharing learning intentions and criteria for success</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Engineering effective classroom discussion and other learning tasks that elicit evidence of student understanding</td>
</tr>
</tbody>
</table>

**Table 2.2 Continued from previous page**
2.1.5 Problems with formative assessment

A noted problem with formative assessment is that its etymology implies a simpler definition than has been discussed in the literature. Sadler (1989, p. 120) noted that “etymology and common usage associate the adjective formative with forming or moulding something, usually to achieve a desired end”. This proves to be unhelpful, since it fosters an assumption that formative assessment is merely an assessment that provides feedback (Bennett, 2011), or frequent assessment, or the product when teaching and assessment are planned at the same time (Assessment Reform Group, 1999). These notions fail to capture the nature of the feedback and how the feedback should be used.

The etymological problem seems to proliferate among lecturers: Taras (2008) conducted a survey with 50 lecturers in an Education department at an English university: of the 48 that replied, only 14 (28%) mentioned feedback when asked for a rough definition of formative assessment.

Such a misunderstanding gives rise to assessment packages that claim to provide formative assessments. Bennett (2011) debated the difference between those that believe that formative assessment is a tool and those that believe that it is a process, concluding that the view that formative assessment is a tool “is quite common among test publishers; it represents something they understand and can provide” (Bennett, 2011, p. 6).

It is the author’s view that this interpretation of formative assessment is incorrect, particularly with regard to the strategies of formative assessment offered by Black and Wiliam (2009). That is, it does not necessarily follow that offering more frequent assessment and more frequent feedback will result in students being made aware of goals and taking actions to progress learning towards these goals. In turn, many assessment practices are assumed to be formative and, therefore, adequate.

The lack of a universal definition not only excuses insufficient practice, but it creates problems when assessing the effects of formative assessment. Bennett (2011, p. 8) noted that “definition is important because if we can’t clearly define an innovation, we can’t meaningfully document its effectiveness”, adding that in determining its effectiveness, it must be known “whether the formative assessment was implemented as intended, which we cannot accomplish if we don’t know what was supposed to be implemented”.

Bennett (2011) also had concerns for the origin of the effect sizes expressed by Black and Wiliam (1998) and by Bloom (1984). Black and Wiliam (1998, p. 3) defined effect size as “the average improvement in pupils’ scores on tests” divided by “the range of scores that are found for typical groups of pupils on these same tests”. Bennett noted that Black and Wiliam did not cite their sources for the data they used and “as such, these
effect sizes are not the ‘quantitative result’, meta-analytical or otherwise, . . . but, rather, a mischaracterisation that has essentially become the educational equivalent of urban legend” (Bennett, 2011, p. 12).

The theoretical underpinning of formative assessment has also been questioned in the literature. Bennett (2011, p. 19) described the background of formative assessment as “not yet [representing] a well-defined set of artefacts or practices”. Taras (2010a) also challenged the theoretical background of formative assessment as proposed by Black and Wiliam (2009), arguing that they had omitted theory of both assessment and formative assessment. Furthermore, challenging the conclusions in an earlier paper by Black and Wiliam (1998), Taras (2010a) suggested that it would not be possible to select the most appropriate papers for analysis without a theoretical basis.

To answer the question “is theory necessary?”, Yorke (2003, p. 484) answered “It is, since theory provides a framework for the construction of assessment of various kinds. Untheorised assessment (as is widely used in higher education) increases the risk of partiality” — as Taras (2010a) had noted.

### 2.1.6 Developing a definition of effective assessment

In order to overcome the problems associated with formative assessment, a framework for evaluating an assessment is necessary: doing so requires developing a definition of effective assessment that encompasses the themes and proposals suggested by existing literature. Such a theory should encompass the tenets of formative assessment; the sharing of responsibilities for assessment; the aims of assessment; and its outcomes. It is of concern that efforts to develop a theory for formative assessment have yielded little by way of a universally accepted framework.

Efforts towards developing such a theoretical framework for formative assessment began with Black (1998), who proposed some elements for such a framework, including:

1. A general learning theory with emphasis on constructivism . . .
2. Models for epistemology of each subject and hence of learning progress within each . . .
3. A theory of the cognitive acts of learning through feedback . . .
4. Analysis of self- and peer-assessment and of the particular learning processes and social interactions these involve.
5. Study of the effects of different styles of feedback on self-esteem, self- attribution and readiness to learn.
Yorke (2003) commented that drawing each of these elements into a single framework is not a straightforward task. Furthermore, he argued that other considerations must also be included:

1. The epistemological structure of the relevant discipline(s);
2. The ontology of students...;
3. Theoretical constructs relating to learning and assessment;
4. The professional knowledge of the educator/assessor...; and
5. Theory relating to communication and interpretation.

(Yorke, 2003, p. 486)

In assembling these various elements, Yorke developed an “illustration of the process of formative assessment in respect of a formal task” (Figure 2.1). It considers the actions and interpretations of both the student and lecturer in the formative assessment process. The process is described as follows.

When a lecturer designs an assessment task, he or she considers the subject structure, the current standard (“general intellectual development”) of the cohort, and “the sequencing
of intellectual and moral development progression” of the student. Lecturers communicate the criteria for the assessment, and the student interprets these criteria as he or she completes the assessment. The lecturer compares the student’s performance with the criteria and interprets this comparison. As a result of this comparison, the lecturer carries out his or her own appraisal of the assessment task, making adjustments to the task or the criteria, while also providing a grade or feedback. The student interprets the grade and feedback leading to further development and the renewing of the process.

It is the author’s opinion that there are some notable absences from this illustration. Yorke (2003, p. 489) later mentioned the issue of “learned dependence”: that students become dependent on lecturers for feedback on their performance without learning how to acquire this knowledge for themselves. The process in Figure 2.1 lacks the scope for students to learn independence and rely less on their lecturers for feedback. It is particularly apparent at the culmination of the “student’s performance” action, where students are not expected to compare their own performances with the assessment criteria.

The second flaw, in the opinion of the author, is the placing of the “Grade, feedback” action. It is implied that students receive feedback only from the assessor and only after the assessment has been completed. Figure 2.1 implies that feedback prepares students for subsequent assessment and it is not clear whether students have the opportunity to demonstrate that feedback has aided their development. Yorke noted:

“If the student has moved to the upper end of the pre-existing zone of proximal development, then he or she should be able to do unaided what previously needed knowledgeable support.” (Yorke, 2003, p. 496)

Thirdly, the formative assessment process, as illustrated in Figure 2.1, does not suggest any role for peers in formative assessment. In both Black and Wiliam (2009) and Nicol and Macfarlane-Dick (2007), there is a responsibility for peers to share learning objectives and dialogue to facilitate improvement. Topping (1998, p. 268) noted that, while more research was desired on the topic, peer assessment has been associated with “improved confidence and better presentation and appraisal skills” and “learning gains in terms of test performance, skill performance, or subjective measures are frequently reported”. The author believes there is a place for peer responsibility in a model for effective assessment.

The model also fails to consider the motives for assessment and how they impact on the criteria. As Black and Wiliam (2009) pointed out, teachers have learning intentions, but students might be working towards their own goals. As with the argument for student autonomy and self-regulation, students must bear some responsibility for their own development and progress towards independence; yet, the lecturer may set a standard that
the entire cohort is expected to achieve. The outcome might be implicit learning outcomes that the lecturer expects to be achieved and some explicit criteria for assessment success that is shared between lecturers, peers and students.

Thus a theoretical framework for effective assessment must incorporate more than what has been highlighted as required for a theory of formative assessment. Such a framework must consider:

- The roles and responsibilities of lecturers, peers and students in achieving effective assessment;
- The motives of lecturers, peers and students and how these affect the criteria for achieving effective assessment;
- The role of practice and feedback in improving the student, particularly with respect to the depth of learning and understanding, and the aspects of students that the feedback is aimed towards;
- The process of developing self-regulation, independence and autonomy in students that should ultimately result in a much-diminished role for the lecturer;
- The constraints that are imposed on assessment that may affect its scope for being effective;
- The influences and requirements that mathematics learning and assessment have on the process of effective assessment.

In the following chapter, the aspects of assessment noted here and previously are discussed in light of activity theory. By using activity theory as a theoretical framework for both the study and the definition of effective assessment, some hypotheses for predicting lecturer, peer and student activity are generated and tested. This is discussed in more detail in the activity theory chapter; however, it is more pertinent to discuss computer-aided assessment and how its use in mathematics assessment appears to relate to effective assessment.

### 2.2 Computer-aided assessment

This section of the literature review shows that computer-aided assessment (CAA) as a tool for assessing mathematics in higher education is increasing in both its uptake and in its complexity.
Although computer-aided assessment (CAA) is by no means limited to mathematics learning (Dermo, 2009; Denton, Madden, Roberts, & Rowe, 2008; McKenna & Bull, 2000; Wilson, Boyd, Chen, & Jamal, 2011; Petrisor, Marusteri, Simpalean, Carasca, & Ghiga, 2016), to higher education (Wang, 2008; Ashton & Wood, 2006), or to the United Kingdom (Chua, 2012; Deutsch, Herrmann, Frese, & Sandholzer, 2012; Rasila, Havola, Majander, & Malinen, 2010), this review of the literature examines papers that refer to the use of CAA systems in higher education for mathematics learning.

However, while it has become much-used assessment tool, offering many benefits in terms of the immediacy of feedback and in saving lecturers’ time, there have been calls for an evaluation of mathematics CAA systems as early as 2003 (Bull and McKenna), even though they have been used for much longer (for example, Roebuck, 1972). Hitherto, evaluations of these systems have largely been a description of new features, rather than their effects on learning.

This section concludes by proposing that the review of a CAA system requires an examination of the impacts the system on learning and development. There has not been a precedent for this kind of evaluation, and given the need for an evaluation framework for the effectiveness of an assessment, this thesis provides an opportunity for a critical assessment of a particular CAA system.

2.2.1 What is computer-aided assessment?

The literature refers to many related phrases that are used in the field: computer-assisted assessment, computer-aided assessment, computer-based assessment and e-assessment are often used and most relevant to this introduction. On first glance, these terms appear to be indistinct in their definitions, as they do not immediately imply any defining differences.

Historically there have been some differences. Computer-aided assessment and computer-assisted assessment (both abbreviated to CAA in the literature) represented all assessments in which information technology has been used as a mediating tool. They include optimal mark readers, where students would complete assessments by hand for computers to scan and mark the responses (Bull & McKenna, 2003; Busuttil-Reynaud & Winkley, 2006). Computer-based assessments (CBAs) delivered questions to and accepted answers from students via a computer terminal (Bull & McKenna, 2003; Higgins & Bligh, 2006; van der Kleij, Eggen, Timmers, & Veldkamp, 2011; Deutsch et al., 2012). An e-assessment adds a further requirement for a networked system of computer terminals to communicate with a server to receive assessments (Bull & McKenna, 2003).
In practice, contemporary systems used in mathematics teaching and learning are e-assessments — they use the Internet to deliver tests and collect responses — but refer to themselves as computer-aided assessments or computer-assisted assessments (Green et al., 2004; Greenhow, Zaczeck, & Kamavi, 2011; Sangwin, 2006; Perfect, Foster, Youd, & Aston, 2011).

This literature review, and subsequently within this thesis, refers to computer-aided assessment since it is most prevalent within the literature. For clarity, the term should be understood in the following way.

**Definition**  A **computer-aided assessment** (CAA) is a test delivered to a computer terminal, where responses are also received from the student. In a CAA system, there is a server in which tests and responses are stored and distributed, which are connected to the computer terminals via a network or, in particular, the Internet.

### 2.2.2 The benefits of computer-aided assessment

Computer-aided assessment offers many benefits. The association between rising uptake of computer-aided assessment and rising student numbers is no coincidence: the automation of assessment promises to save lecturers’ time (Bull & Stephens, 1999; Thelwall, 2000) and give students immediate feedback (Bull & Stephens, 1999; Pollock, 2002), where alternative forms of assessment may require additional markers or less timely feedback. The perceived benefits of computer-aided assessment go further.

Bull and McKenna (2003) gave seven reasons for using computer-aided assessment:

1. To increase the frequency of assessment thereby motivating students to learn encouraging students to practice skills
2. To broaden the range of knowledge assessed
3. To increase feedback to students and lecturers
4. To extend the range of assessment methods
5. To increase objectivity and consistency
6. To decrease marking loads
7. To aid administrative efficiency

(Bull & McKenna, 2003, p. 8)

Since these seven motivations for using CAA cover many aspects of teaching, learning and administration, it is worthwhile exploring each of these points in more detail and how they are discussed in the literature.
Increasing the frequency of assessment

Green et al. (2004, p. 5–6) argued that computer-aided assessment, when delivered regularly, benefits students in a number of ways. Students become familiar with the system and are less agitated by stress. Furthermore, regular testing permits students to monitor their own progress at the end of each taught topic; each test represents the culmination of study in an area of mathematics, bringing structure to the module.

Whether an increase in assessment frequency brings greater motivation in students is not clear from the literature. Students appear to be motivated by the delivery of immediate feedback (Beevers, Wild, McGuire, Fiddes, & Youngson, 1999; Denton et al., 2008); however, this motivation may be achievement-based and not directed towards learning and understanding mathematics (Thelwall, 2000). Indeed, Bull and McKenna (2003, p. 10) concede that students “are reluctant to undertake work which does not count towards their final grade”.

Although increasing the frequency of assessment is not the sole encouragement students receive, Harrison, Green, Pidcock, and Palipana (2007) indicated that students were encouraged by a regime of regular assessment. Sangwin (2007, p. 1) made a similar point: “It is relatively common in United Kingdom undergraduate mathematics to set weekly exercise sheets. While the original purpose is formative, to encourage student engagement the numerical mark provides a small summative contribution to the overall course”.

Broadening the range of knowledge assessed

In expanding upon their earlier claim that one reason for using computer-aided assessment is to broaden the range of knowledge assessed, Bull and McKenna implied that computer-aided assessment cannot be used in isolation to assess students.

“Objective tests [such as computer-aided assessment] lend themselves well to testing a broad knowledge base within a particular discipline. In many courses, there is a body of underpinning knowledge which must be learned to enable progression during the later stages of the course... Different assessment methods are effective at assessing different skills, knowledge and understanding. ...It is clearly important that a balance of assessment methods is used in order to assess the full range of skills and knowledge required in courses.” (Bull & McKenna, 2003, p. 10–11)
In this passage, Bull and McKenna do not explain the role of computer-aided assessment in assessing the “full range of skills and knowledge”. It is the author’s opinion that Bull and McKenna have implied that computer-aided assessment was not able to bear the full burden of assessing this “full range” at that time: rather, computer-aided assessment could be used effectively to assess students’ lower levels of understanding, while freeing other assessment methods to test more complex skills and deeper knowledge.

Beevers and Paterson (2003, p. 132) suggested this may be the case: “It is far from certain, however, that objective tests [such as computer-aided assessment] can assess learning beyond basic understanding. However, questions that are constructed imaginatively can challenge students and may test higher learning levels”.

The design and the capabilities of aspects of the computer-aided assessment system can be limiting factors in the depth of knowledge that can be assessed. Conole and Warburton (2005) summarise the debate:

> “The limitations of item types are an ongoing issue. A major concern related to the nature of objective tests is whether multiple choice questions (MCQs) are really suitable for assessing higher-order learning outcomes in higher education students... and this is reflected in the opinions of both academics and quality assurance staff... The most optimistic view is that item-based testing may be appropriate for examining the full range of learning outcomes... provided sufficient care is taken in their construction.” (Conole & Warburton, 2005, p. 21, emphasis in original)

This echoed the conclusion in their earlier paper (Warburton & Conole, 2003, p. 436), in which they stated, “The findings suggest that typically quiz questions are considered limited in their capacity to assess higher-order cognitive skills.” However, Nicol (2007, p. 62) argued that multiple choice questioning has a place in assessment when linked to “a clear pedagogical goal”.

### Increasing feedback to students and lecturers

As a motivation for using computer-aided assessment, Bull and McKenna (2003, p. 11) proposed that “students are motivated by feedback on their work”. Although they added that feedback needs to be “timely”, “accurate” and “constructive”, they did not elaborate further to explain how increasing the feedback students receive is cause to motivate them.

The notion that students, given suitable feedback, have scope to improve their knowledge and understanding is more generally accepted. Measuring the effect of this feedback is
particularly troublesome, especially in higher education mathematics. Within computer-aided assessment, the situation appears no better: van der Kleij et al. (2011) performed a small-scale study exploring the effect of feedback on students’ learning outcomes while using computer-based assessment and concluded that there is little evidence to suggest this feedback aids students’ performances. Although a small study, it serves to demonstrate that merely providing feedback does not necessarily make the exercise effective in helping students to improve.

**Extending the range of assessment methods**

There is a reliance on so-called “traditional” methods to assess students (Iannone & Simpson, 2012), yet computer-aided assessment offers the scope for multimedia presentations beyond the capability of paper-and-pencil assessment (Bull & McKenna, 2003).

Thelwall recommends a blend of traditional and computer-based assessments:

> “Computerisation of assessment can enhance the value of education in certain circumstances, either because it is intrinsically better than paper assessment or because the assessment would be impractical without it. Computers cannot be used for all assessments because some educational outcomes, particularly those with a creative component, are too difficult to measure with the defined rule that a computer program needs. The point here is that CBA is different to and, in the information age, complementary to traditional assessment methods.” (Thelwall, 2000, p. 47, emphasis in original)

**Increasing objectivity and consistency**

Bull and McKenna (2003) highlighted the growing importance of objectivity and consistency in a higher education market in which students are assuming the role of customer. Computer-aided assessment offers transparent marking (Conole & Warburton, 2005), which is apparently objective and consistent, since it removes human subjectivity and accuracy that are present in traditional marking.

**Decreasing marking loads**

Peat and Franklin (2002) posed the challenge faced by lecturers ten years ago, albeit in an Australian university in the biological sciences: in the ten years leading to 2002, the staff-to-student ratio has changed from 1:100 to 1:163. For STEM subjects in the United
The increase in student numbers is similarly stark: the number of students graduating has increased from approximately 118,000 in 2002–2003 to over 140,000 in 2009–2010 (Select Committee on Science and Technology, 2012, p. 27). In 2013–2014, there were 1,007,630 students in STEM subjects at UK universities (Universities UK, 2015). Growing student numbers would naturally increase the number of assessments that would need to be marked.

Computer-aided assessment provides a mediating solution: though it cannot eradicate the marking required of lecturers, it permits lecturers to set assessments with automatic feedback (Charman, 1999; Bull & Stephens, 1999; McKenna & Bull, 2000; Peat & Franklin, 2002). However, as Bull and McKenna (2003) noted, the time saved in distributing and marking assessments must be considered alongside the time expended in implementing and updating the system. This in turn can be offset if the maintenance of the computer-aided assessment system is performed by non-teaching administrators.

Aiding administrative efficiency

Computer-aided assessment can increase administrative efficiency by collating students’ scores automatically (Bull & McKenna, 2003). CAA also manages the distribution of tests, the collection of responses, marking the responses and distributing feedback. For more traditional forms of assessment, these tasks take considerable time.

2.2.3 The use of computer-aided assessment in higher education mathematics learning in the United Kingdom

Since the turn of the century, many papers have reported that CAA use is increasing steadily (Denton et al., 2008; Keady, FitzGerald, Gamble, & Sangwin, 2006; McKenna & Bull, 2000; Özden, Ertürk, & Sanli, 2004; Pitcher, Goldfinch, & Beevers, 2002; Thelwall, 2000) and this has been attributed to the large increase of students, particularly in STEM subjects (Bull & Stephens, 1999; Davies, 2001; Gill & Greenhow, 2008; Jones, 2008; Krause, Stark, & Mandl, 2009). This claim seems to be no longer repeated in new articles and books on computer-aided assessment: development has been focused on developing the complexity and capability of those systems, rather than the spread of practice.

Miller (2009, p. 183) noted that “research on formative CBAs has focused on the development and evaluation of these assessments”, largely from the lecturers’ perspectives. These contain some details of the location and the system that is used, but few indicate the scale of the deployment in terms of cohort size or number of tests taken. There lacks
a broad study of the current uptake of CAA in mathematics courses in Higher Education in the United Kingdom.

A previous study examined the use of CAA to perform mathematics diagnostic tests. In its report, “Diagnostic Testing for Mathematics”, the LTSN MathsTEAM Project (2002, p. 6) reported that 36% of the institutions polled used such testing. This only gives an indication of the uptake of CAA: the study was performed more than ten years ago. However, the study reported that there was already a number of different systems (and systems derived from those systems) in use that are now obsolete, or have been replaced by newer systems.

This demonstrates some of the difficulty in presenting the nature of CAA use in higher education mathematics in the United Kingdom. From the literature it is apparent, yet not transparent, that the adoption of CAA is increasing. Confounding a unified understanding of the current situation is the confusion between different terms and a difficulty in differentiating between system types.

There are, however, some studies that report lecturers’ perspectives of implemented computer-aided assessment systems in mathematics higher education courses in the United Kingdom (Bull & McKenna, 2003; Greenhow et al., 2011; Jones, 2008; Pidcock, Palipana, & Green, 2004; Pitcher et al., 2002; Pollock, 2002; Ricketts & Wilks, 2002; Sangwin, 2005, 2006, 2007). Studies of this nature are useful in identifying issues in computer-aided assessment—whether confined to mathematics at higher education level or otherwise—but they often lack the student-user perspective (Walker, Topping, & Rodrigues, 2008). Some student feedback studies of computer-aided assessment have been conducted in response (Denton et al., 2008; Walker et al., 2008), but little progress has been made in such regard for mathematics computer-aided assessment.

A notable obstacle is that mathematics assessment can be regarded as a special case. One might take for granted that communicating mathematics verbally between humans is difficult. Written communication of mathematics may seem easier, but simple mistakes can proliferate. Communicating mathematics with a computer proves to be a hybrid affair, where a human must provide a “one-dimensional string” of the correct syntax in order to successfully convey the appropriate message (Sangwin & Ramsden, 2007, p. 921).

There are two solutions to this problem: either incorporate a computer algebra system (CAS) that interprets mathematical input and offers students the opportunity to review the rendering of their responses before submitting for marking; or adopt a system that does not require students to learn a syntax. The systems currently available are divided into these two categories.
Computer-aided assessment supported by a computer algebra system

With the support of a computer algebra system (CAS), a CAA system can compare a mathematical expression, provided by the student, with the stored solution for equivalence. That is, a student may obtain full marks for answering a question correctly without matching the stored solution verbatim (Sangwin, 2004, p. 5).

The computer algebra system determine any differences between the student’s response and the stored solution. If there is no difference, the appropriate marks are awarded. If there is a difference, the student may be awarded partial marks if it appears he or she has made a common error (Sangwin, 2004).

Such tolerances are particularly useful when the solution is algebraic in nature, given that the presentation of mathematical expressions are sometimes a matter of personal style: for example, $-x^2 + 1$, $1 - x^2$ and $(1 + x)(1 - x)$ are equivalent but, stylistically, they are different. Additionally, there may be situations in which one or more of those solutions is not acceptable, such as if the question asks to expand $(1 + x)(1 - x)$, then clearly the last expression of the three would not be desired; in that case, such solutions can be excluded from the set of expressions that are marked correct (Sangwin, 2007).

Once a question has been designed, for many CAA systems it is relatively straightforward to introduce elements of randomisation. For example, “genuinely random” polynomials of degree and coefficients within predetermined ranges can be generated (Sangwin, 2004, p. 5). This has two advantages: students can practice performing differentiation on polynomials repeatedly; and it is also unlikely for two students sitting side-by-side to receive identical realisations of these randomisations, offering a barrier to plagiarism. This randomisation does not require question-writers to manually input every possible instance: it can be achieved by requesting the CAS to generate the question instance instantaneously.

Sangwin (2004) demonstrated how this is useful for generating feedback for students. In the case of integrating a function, the computer algebra system is able to differentiate the student’s response and present this to the student. This way, the student is not given the solution, but is offered a reason why his or her answer is not correct and an indication of what corrections might be required.

At the time of his conference paper, Sangwin (2007) mentioned some CAA systems that were supported by a number of different computer algebra systems. This may be a problem since students not only need to learn the material, they also need to learn the syntax. Sangwin and Ramsden (2007) noted this problem with regard to AiM. One student had commented in feedback, “I feel the aim (sic) system is reasonably fair,
however i (sic) have lost a lot of marks in quiz 3 for simple syntax errors” (Sangwin & Ramsden, 2007, p. 921).

Although the CAS can indicate where it cannot understand a human response (Sangwin, 2002), it does not necessarily follow that the CAS will interpret a student’s response correctly (Sangwin & Ramsden, 2007). Clearly, such a result is “unacceptable” (Sangwin & Ramsden, 2007, p. 922): the assessment becomes as much a measure of a student’s ability to converse fluently with the CAS as it is a measure of a student’s mathematical performance.

A further criticism of a CAS-supported CAA system is that there may be a financial outlay for the CAS software, even if the CAA itself is free. For example, AiM is free, but requires the purchase of a single Maple licence (Keady et al., 2006). Others, however, require no financial outlay for software, such as Numbas and STACK, but may require investment in new hardware (Keady et al., 2006).

Computer-aided assessment without a computer algebra system

For CAA systems that lack an ability to interpret mathematical input, asking questions of a mathematical nature is a challenge. Many pen-and-paper assignments ask students to answer questions that result in an algebraic (or ‘mathematical’) form that has no clear analogue when performed electronically. Consequently, questions must be reconsidered before they are generated for the CAA system (Conole & Warburton, 2005).

CAA systems that are not supported by CAS cannot interpret an algebraic response. They may use other question techniques, such as multiple choice, multiple response, text or numeric input, matching, ranking and drag and drop (Conole & Warburton, 2005; Bull & McKenna, 2003). Sangwin (2004, p. 5) argued that converting questions to these alternative forms “often limit or distort questions”.

There are many CAA systems used for mathematics assessment that do not rely on a CAS (Conole & Warburton, 2005). Perception, by Question Mark Computing Ltd, is a popular choice (Dermo, 2009; Green et al., 2004; Greenhow & Gill, 2005; van der Kleij et al., 2011; Martin & Greenhow, 2004; Walker et al., 2008). This commercial choice offers more test options but requires considerable initial outlay (Conole & Warburton, 2005).

Although there are several options for question types and systems, multiple choice questioning remains the prominent choice for many CAA systems (Warburton & Conole, 2003). Whether multiple choice questioning (MCQ) is appropriate for testing knowledge and understanding is an issue that is contested considerably within the literature.
Bull and Stephens (1999) referred to an earlier study performed by Scouller and Prosser (1994) in which they found that students approached multiple choice questioning differently: students with deep learning intentions were more likely to adopt study strategies for MCQ examinations; students with surface learning approaches tended not to approach MCQ examinations with a study strategy, and furthermore they were more likely to not fully understand the difference between “understanding” and “reproducing” factual knowledge.

Butler, Karpicke, and Roediger (2007) discussed ways in which multiple choice questioning could be improved. They found that there was no significant difference between students’ long-term retention of knowledge whether they received “standard” feedback (marking correct or incorrect and giving the solution) or “answer-until-correct” feedback (marking correct or incorrect and giving the student another attempt if they answered incorrectly until he or she answers correctly). Butler et al. (2007) concluded that delaying feedback leads to better performance, indicating an improvement in long-term retention.

Bush (2001) gave further assurances that multiple choice questioning can be effective if sufficient care is used in writing questions. This effect may be enhanced with negative marking, with enabling the student to give more than one response, or with students’ self-assessment of their confidence in their solutions. He added a note of caution that such testing can cause confusion, particularly if they are not used universally.

Greenhow and Gill (2005) explained that the success of multiple choice questioning is aided by analysis of previous cohorts’ responses to similar questions. By establishing common errors that have been made by students previously, the lecturer is better able to determine the root of the mistakes. In doing so, multiple choice options can include “mal-rules”: answers generated “if the student applies sensible, but incorrect, rules [methods] of their own” (Greenhow & Gill, 2005, p. 2). Furthermore, by anticipating those common errors, the lecturer can write feedback explaining the error and how this can be avoided in future. For some CAA systems, this feedback can be generated as simply and as quickly as it can generate the questions — so long as the system has been informed how to generate this type of response. Without awareness and input of the mal-rules, the CAA system is not likely to know where the errors have occurred and feedback to such detail cannot be given.

Some feel that CAA systems that rely on multiple choice questions and numerical input — which some refer to as “quizzes” — cannot test deeper understanding and cannot be relied upon for summative assessment. Indeed, both lecturers and students realise this: “[It] can be argued that such CAA tests do not adequately assess the higher level skills or understanding in mathematics. To a large extent this is true. Students are well aware of this, distinguishing as they do in feedback, between credit gained for method
and credit gained for a single numerical input” (Croft, Danson, Dawson, & Ward, 2001, p. 65, emphases in original). Conole and Warburton (2005, p. 21) agreed: “A major concern...is whether multiple choice questions are really suitable for assessing higher-order learning outcomes in higher education students”.

Although CAA systems, with or without a supporting CAS, offer much to both lecturers and students, it is apparent that they bring some issues. The following section highlights the persistence of these problems.

### 2.2.4 The problems with computer-aided assessment

There are many professional papers that extol the benefits of computer-aided assessment. These papers tend to report the experiences of a lecturer implementing a CAA system. There are few longitudinal studies that examine the long-term implications of the installation and on-going use of CAA, yet there are some reviews of CAA that attempt to collate the issues. The following discussion is a fresh attempt to achieve a similar result.

To classify and explain the relationships between issues, this discussion is divided into three categories, as chosen by McKenna and Bull (2000, p. 25): pedagogical, operational and institutional. Although they do not operate as exclusive or distinct categories — they are “clearly inter-related” — there are internal links within categories that provide much scope for discussion.

**Pedagogical issues**

Perhaps the most pressing issue for CAA is ensuring that it is accepted by lecturers. There appear to be a number of reasons why lecturers would rather not use CAA with their students. Some feel that CAA is part of a process that makes education impersonal: “It is fair to say that a concern of tutors was that automated feedback was another step on the road to an uncertain impersonal future” (Barker & Lilley, 2006, p. 51). Likewise, some lecturers prefer to have an active role in the feedback process, as Barker and Lilley explained:

“Tutors liked the ability to be able to ‘keep a finger on the pulse’ when providing feedback. Some concern was expressed that an automated approach would lead to potential problems going unnoticed. This could not happen when tutors themselves gave feedback.” (Barker & Lilley, 2006, p. 49)
The design of questions has already been identified as an issue with respect to assessing understanding and knowledge. Writing questions that adequately meet this requirement was a concern raised by Stephens and Curtis:

“Previous experience of question writing was judged to be mixed amongst lecturers, which can result in questions being provided by lecturers that may not always tie in with their course objectives and may not be assessing at the appropriate level of learning.” (Stephens & Curtis, 2002, p. 19)

In addition, there are limitations to the scope of what can be tested using CAA at the current time: assessing proof and reasoning skills using CAA is “difficult to envisage” (Sangwin, 2004, p. 7). Nonetheless, McKenna and Bull believed that assessing deeper understanding can be achieved, but indicated that it requires more thought and creativity:

“The capacity of objective tests to assess a wide range of learning is often underestimated. Objective tests are very good at examining recall of facts, knowledge and application of terms, and posing questions that require short text or numerical responses. Additionally, questions, which are constructed imaginatively, can test analysis, synthesis and [evaluation].” (McKenna & Bull, 2000, p. 26)

Though objective tests appear transparently fair, Dermo (2009, p. 211) noted that the selection of questions at random from a question bank was not universally well-received — “clearly there is a perception among students that these are unfair” — and ensuring similar questions are of equal difficulty is another burden that question designers must bear.

Some students also feel that, with some CAA marking schemes, they do not receive the marks they deserve. For many systems, marks are awarded only for final answers since students cannot communicate their method to the computer. Consequently, “many students felt uneasy with numerical input questions, feeling that too many marks were lost if they made slight numerical errors” (Croft et al., 2001, p. 65).

Furthermore, there is an added expectation that students are comfortable using information and communication technology to complete CAA tests. Although the study did not involve mathematics assessment, a study performed by Biscomb, Devonport, and Lane suggested that this might be a problem: “one of the major disadvantages noted by staff was the issues of students’ levels of IT competence” (Biscomb et al., 2008, p. 87).
The lack of reward for marking and the assumption that students are capable of using the technology may be regarded as both operational and pedagogic issues. While they are limitations of the operational capabilities of the system, they are also restrictions to the pedagogical decisions that lecturers must make with regard to assessment.

**Operational issues**

McKenna and Bull (2000, p. 28) viewed operational issues as being barriers to performing the assessment to ‘rigorous exam conditions’. However, it is not always the intention for assessment to have the same purposes as exam conditions. If CAA is to be used with the intention to provide feedback to the student, then limitations in what the CAA system can provide in terms of feedback becomes an operational issue.

Operational issues may also have more widespread effects. The first issue that McKenna and Bull raised is that of timetabling: that is, CAA tests ought to be timetabled so as to minimise the chance of conferring. As they noted, there are instances where a cohort is too large to be accommodated in a single computer laboratory (Beevers et al., 1999; Conole & Warburton, 2005). In such instances, McKenna and Bull advise testing one group immediately after testing the other in order to ensure no conferring is possible.

Regardless whether invigilation is possible, many CAA systems feature a randomisation technique that makes it highly probable that no two students in a cohort will receive identical tests (Gikandi, Davis, & Morrow, 2011; Green et al., 2004; Pollock, 2002). By using this technique, students are faced with an obstacle to cheating, although it does not eliminate this cheating completely (Green et al., 2004). The effect of this cheating is unknown — “we really have little idea of the true extent of this problem outside invigilated labs” (Sangwin, 2005, p. 6).

The issue of security was another concern of McKenna and Bull: primarily the storage of the question and answer files, and students’ responses; but also ensuring that students’ identities are checked as they take CAA tests. As for storage, McKenna and Bull (2000, p. 28) reported that, at the University of Luton, question and answer files are removed from the server when they are not in use.

A further point is that technology is not infallible (McKenna & Bull, 2000). Ke, Yingwei, Yajun, and Runhua (2010) added:

“Computer equipment has its own limitations, it may not always be available, or be in reliable working order. Additionally, test takers’ mental status, such as scare [fear] and anxiety, has a negative impact on tests. Screen size, and
graphics capabilities could also be limitations in practice of E-assessment."
(Ke et al., 2010, p. 1659)

Some equipment and software issues affect some students more so than others when they occur. Biscomb et al. (2008, p. 87) noted that one of the lecturers “considered that the use of technology might be more problematic for students with special educational needs, with particular reference to dyslexic students”.

This is an aspect of mathematics-based CAA that has been considered by Greenhow (2000), with the use of font, size and colour to aid visually impaired students, and by Sangwin (2005), with the implementation of mathematics rendering that can be used with students that are visually impaired. Though such adjustments have been introduced to aid those with visual impairments, it is not clear from the literature the extent to which other disabilities have been considered.

Time appears to be a limiting factor in the maintenance and updating of a CAA system, particularly when considering the development of new CAA tasks. Designing questions has been identified as a time-consuming process that warrants careful consideration. It appears in reality that lecturers seldom have the necessary time for such an undertaking (Stephens & Curtis, 2002).

Students must also manage their time using CAA around other commitments: having access to the tests at all times can provide some relief to their timetables (Pollock, 2002). However, students desire support during assessment and the 24-hour availability of CAA can lead some students to believe that such support is also available all day (Bertolo & Lambert, 2007).

**Institutional issues**

McKenna and Bull (2000) believed that having adequate staff development to ensure that novice and experienced practitioners could maintain and update the system was important for the ongoing success of the system.

It could be argued that a lack of such training is a missed opportunity for the discussion of good practice. With many options available, lecturers approach CAA “in an individualistic manner. (Stephens & Curtis, 2002, p. 19). This might cause some confusion for students that experience different approaches from different lecturers. McKenna and Bull (2000, p. 27) reinforced this point: “Stephens, Bull, and Wade (1998) have argued for a systematic, fully-embedded implementation of CAA across an institution and suggest nine recommendations for its successful introduction”. There is little by means of
the governance of CAA use within higher education institutions; however, this seems no different from the governance of assessment more generally:

“Of the 25-quality assurance staff who returned the CAA Centre questionnaire, only one reported having separate exam regulations in place for the governance of CAA and this was specific to the use of OMR in formative coursework, not for summative CAA. This finding is consistent with Yorke’s observation that managing assessment at an institutional level is an underdeveloped practice in UK higher education.” (Yorke, 1998).

(McKenna & Bull, 2000, p. 27)

Stephens et al. (1998) placed great importance on the cohesion of staff members to ensure that CAA becomes a success:

“It is essential to have the support of all staff involved in designing, implementing and maintaining such systems, including academic, technical and support staff. Suspicion surrounding new initiatives, innovations and technology makes it crucial that every effort is made to ensure the smooth running and effective operation of such systems. This can only be achieved through a strategic mandate with general institutional approval.” (Stephens et al., 1998, p. 285)

Achieving universal support for CAA, however, seems challenging. Some lecturers — and students — might prefer to conduct tests on paper; others feel that CAA does not adequately test students. Consequently, there may be much support for CAA while some remain sceptical:

“The introduction of CAA in the form of objective tests in multiple choice question format has generated much but not universal enthusiasm among students and staff.” (Woodbury, Ratcliffe, & Thomas, 2001, p. 16)

2.2.5 Summary

It is clear that many issues are present in the implementation of computer-aided assessment; in the case of mathematics assessment, there are particular issues that are unique to communicating and assessing elements of the syllabus. However, we remain interested in the effects of CAA in terms of student improvement. For this, we turn to the practice facility available in many CAA systems.
Many papers claim that the CAA system being used offers formative assessment (Bull & Stephens, 1999; Cerval-Peña, 2008; Crisp & Ward, 2008; Denton et al., 2008; Higgins & Bligh, 2006; Sangwin, 2007; Thelwall, 2000), largely because they offer a practice facility. However, there is little evidence in these papers to suggest that the feedback (or “evidence about student achievement”) is routinely “elicited, interpreted and used by teachers, learners, or their peers, to make decisions about the next steps in instruction” (Black & Wiliam, 2009, p. 9, emphasis added).

It has already been claimed that students are motivated by practice and feedback (Bull & McKenna, 2003; Gikandi et al., 2011; Crisp & Ward, 2008) and there is some evidence in the literature that students earn improved scores with repeated practice (Thelwall, 2000). However, van der Kleij et al. (2011) found no significant improvement regardless of whether “elaborated feedback” was given or not.

2.3 Conclusions

Assessment in mathematics courses in higher education is dominated by examinations. There are many reasons for this: they are regarded as fair, valid and reliable. They are also particularly well-established and there are institutional pressures to maintain this practice.

While this type of assessment is the dominant summative assessment tool, CAA is becoming more popular: particularly for interim assessments. CAA offers lecturers the opportunity to save time in delivering and marking assessment and to offer immediate feedback.

There are a number of different systems available and they can be divided into two categories: those supported by a computer algebra system (CAS) and those that are not. Those with a CAS permit a wider variety of mathematics questions and response types: notably this includes algebraic responses, which systems without a CAS cannot offer. The lack of a CAS restricts the question types available: primarily to multiple choice questions and numerical responses.

There are a number of problems that have already been identified with using CAA. Practice is often uncoordinated and disjointed between colleagues; updating systems with new questions is tedious; and it is unclear to what extent students learn through using this type of assessment.
While it has been asserted that practice is important in mathematics, it has also been noted that many students take cues from their assessments when deciding on their approach to learning. Some are satisfied with surface learning, or with adapting their learning to maximise their marks in the assessment.

Therefore, for CAA to be effective, it must at least encourage students to engage in deeper learning; however, the literature provides no evidence for or against this claim.

Many CAA systems are described as formative assessment tools in the literature, but it is not clear — either in the CAA literature or in the formative assessment literature — what this means in practice.

An effective assessment should also require that students improve. This requirement is only sometimes made of formative assessments. It is not apparent from the CAA literature that students make a measurable improvement.

In order to evaluate the CAA system in this study, there must be a clearer definition of what it means for an assessment to be effective. In order to do so, the following chapter presents a theoretical framework from which to view the activity of assessment and, from this position, it is possible to determine a model for what makes an assessment effective. Using this model, one can evaluate an assessment tool — in this case, a CAA system — for its effectiveness.
Chapter 3

Theoretical Framework

Activity theory is a framework for describing and interpreting socially mediated, yet individual, activity: Barab, Evans, and Baek (2004, pp. 199–200) described activity theory as a “psychological and multidisciplinary theory with a naturalistic emphasis that offers a framework for describing activity and provides a set of perspectives on practice that interlink individual and social levels”.

Assessment is a social activity between someone that is assessed — usually a student — and an assessor — usually a teacher. However, assessments have rules that are associated with them; they may be developed as a social enterprise. Cultural-historical activity theory (CHAT) focuses on these aspects of human activity to identify influences, contradictions, outcomes and changes.

This chapter provides an introduction to activity theory and provides a rationale for using CHAT as a theoretical framework for this study. It culminates in a description of how the author used CHAT and the assessment literature to develop a model for effective assessment.

3.1 What is activity theory?

Activity is the most fundamental concept of activity theory. Leont’ev (1978) defines it as any purposeful interaction between a subject and an object. It is activity that is studied in activity theory: Leont’ev described activity as the most basic unit of life; Kaptelinin and Nardi (2006) asserted, citing Leont’ev (1978), that subjects and objects have no discernible properties when there is no activity. Therefore, without studying activity, one cannot deduce properties of the subject or the object at that point. As such, Kaptelinin and Nardi suggested that while activity theory does not preclude one
from speculating properties of a subject or an object based on generalisations, it asserts that these speculations may be inaccurate.

In addition, cultural-historical activity theory considers the historicity of the society and the mediating tools of an activity. For example, the desire to rank and grade students on their ability developed over time; and as such, exams are firmly established as the primary means of assessing students — particularly in mathematics (Iannone & Simpson, 2011).

Yet the individual has their own history that they bring to the activity. Thus while an exam offers a seemingly objective assessment of students’ performance and accomplishment, students that are accustomed to this assessment practice, or have additional experience with the content of the exam, have additional tools from which to draw upon for this exam.

Therefore, it does not follow that the activity of taking a mathematics examination can be described merely by noting the characteristics of the student and the exam. Therefore, it is necessary to analyse the activity, as opposed to examining the subject or the object in isolation.

### 3.1.1 The structure of activity

Not every thing can be a subject: a subject must have needs that can only be satisfied by acting in the world (Kaptelinin & Nardi, 2006). This precludes non-living items from being considered as subjects.

When an object presents itself as an opportunity to satisfy the needs of a subject, it becomes a motive. Although a subject may be driven by a motive to satisfy a need, the subject may be compelled to engage in other activities that do not correspond directly to this motive. For example, a student (the subject) needing to take a CAA test (the object) must switch the computer on (an action). This task, in itself, does not directly lead the subject to satisfy the need; however, this is an important component of the overarching activity.
It follows that activity can be considered as a hierarchy of tasks (figure 3.1). While activity is driven by a motive, actions are driven by goals. Actions are the chief component of activity: switching the computer on is considered an action. Goals are conscious aims of actions: subjects need not be conscious of their motives.

Actions may be further composed of operations, which are lower-level tasks that are automatically carried out without an acute awareness from the subject. Operations are carried out to satisfy conditions: these conditions relate to the goals of the ‘parent’ action. For example, a student may have to scroll down one of the test pages in order to see the rest of a question, without consciously being aware of doing so.

Items within this hierarchy may be promoted or demoted. For example, once the subject is familiar with logging into the CAA system, the subject will log onto the system without investing much thought: thereby this action becomes an operation. Conversely, if the CAA system was redesigned and the process for logging in changed, the student would not be able to perform the procedure without thought and would thus become an action again.

Humans may use tools, known as artefacts, to attain goals. For example, a student may use a bookmark to access the tests. The student needs to know how to use the bookmark (the tool-related competency) and the fact that bookmarks are used to get to websites quickly (the task-related competency).

### 3.1.2 Key ideas within activity theory

Two central tenets dominate activity theory: that consciousness and activity are inseparable; and that the mind is social in nature (Kaptelinin & Nardi, 2006). These ideas mean that the mind exists and can be understood only when activity is taking place, and that society and culture directly influence products of the mind.

The processes of internalisation and externalisation are prominent features of activity theory. It asserts that an activity that once required the use of external tools may be internalised, no longer requiring those tools. For example, students typically learn to perform long division on paper before learning to carry out such calculations mentally. However, the process may reverse; students may resort to using a calculator if the solution seems incorrect: this is an example of externalisation, where a previously internal action requires the assistance of a tool. Vygotsky described a similar process by which a student, or child, would develop from a social experience before internalising the experience:

> “An interpersonal process is transformed into an intrapersonal one. Every function in the child’s cultural development appears twice: first, on the social
Internalisation also describes the transforming of an activity from a collective, social activity to an individual one. For example, a novice surgeon may require more experienced staff for assistance until the skill and confidence to perform these actions alone have been acquired. Externalisation is the reverse process: the surgeon may require support from others if complications arise in a surgical procedure. That is, interpsychological mental functions — distributed between many people — may become intrapsychological mental functions, and vice versa.

Mediation considers the impact of artefacts on human activity and the culture and history developed by these artefacts. Computers have their own culture and history, having been developed to be faster, more reliable and more efficient. They have significantly altered many aspects of society in culture: in manufacturing, communication and productivity.

Cultural-historical activity theory examines a large number of intricacies and considerations in an activity system. Representing mediation is particularly troublesome, as it applies equally to the subject and its community, and it imposes on the nature of the object. Engeström (1999) introduced a diagrammatical representation of an action that encompasses the relationships between the subject, object and community (figure 3.2).

Engeström (2000) also described “disturbances” as occurrences in an activity that cause tensions. They may lead to contradictions, which are instabilities in an activity system that need to be addressed. They are the drivers of change in activity.

Cultural-historical activity theory provides a framework for human activity and guides researchers to aspects of an activity system to observe for analysis. Activity is social and individuals engage with and negotiate within a community in order to achieve personal...
success. When engaging in a group activity, such as with studying at a university, success is conferred upon an individual by the community; at a university, this is usually in the form of grades, marks and awards. This requires the individual to align with the expectations of the community: that is, the goals of the individual must be compatible with the goals of the community.

These positions guide researchers towards examining the goals of the individual and the community and to what extent these goals align and differ, in addition to how the individuals and the community act towards their goals. With regard to this study into how lecturers and students use a computer-aided assessment, they raise four questions.

• Developing an identity: What effect do students’ and lecturers’ histories and cultures have on the way they use the CAA system?

• How goals are formed: What kind of goals are students and lecturers interested in pursuing, and what happens when they achieve them?

• The influence of the community: What effect does the community have on students and lecturers using the CAA system?

• Resolving contradictions: What happens when students and lecturers encounter a contradiction while using CAA?

The following section explores these aspects in more detail.

3.2 Detailed examination of aspects of CHAT

3.2.1 Developing an identity

The notion of identity is a topic of debate in socio-cultural studies. In a general terms, identity is the set of characteristics that an individual conceives of themselves that express individuality or alignment with a social or cultural group. That is, an individual may consider his gender, age, religion, nationality, sport affiliations and profession to be elements of his or her identity.

Roth explained how cultural-historical activity theory considers identity to be interpreted through activity:

“The construct of identity pertains to who someone is. However, we do not know who a person is independent of the actions of that person. Being shown
the middle finger by another car driver, we may think, “he is a rude person;” overhearing someone using the four-letter f-word incessantly, we may think, “she is crude;” and seeing a group of students performing incredible stunts on their skateboards, we may think “they are skateboard experts.” In all of these examples, attributions about who someone is are made based on observable behaviour (actions). Actions that are already means of expressing emotions and motivations also come to express identities.” (Roth, 2007a, p. 56, emphases in original)

This is certainly true to Leont’ev’s description of activity and consciousness. He believed that in order to understand both the subject and the object of the activity, the researcher must analyse the activity. In particular, Leont’ev (1977, p. 5) noted that in order to understand the internal processes of the subject, “it is external activity that unlocks the circle of internal mental processes, that opens it up to the objective world”. Kaptelinin and Nardi added:

“...activity theory maintains that no properties of the subject and the object exist before and beyond activities (e.g. Leont’ev 1978). These properties do not just manifest themselves in various circumstances; they truly exist only in activities, when being enacted.” (Kaptelinin & Nardi, 2006, p. 31)

Holland, Skinner, Lachicotte Jr, and Cain described identity as something that is constructed by individuals that both describe and influence how they act:

“People tell others who they are, but even more important, they tell themselves and then try to act as though they are who they say they are. These self-understandings, especially those with strong emotional resonance for the teller, are what we refer to as identities.” (Holland et al., 1998b, p. 3)

While individuals construct and develop their own identities, Holland et al. explained that they are subject to influences that might act as constraints. There are social norms and practices, cultural resources and past experiences that must be considered. Consequently, they argued that individuals do not truly act spontaneously: the extent to which their activity is novel is bound by these factors.

Identity is by no means rigid, yet there are limitations to its change. Indeed, an individual may develop multiple identities that are used in different social or cultural situations, and different activities: “one ‘becomes’ what one ‘does’...through reflexivity” (Black et al., 2009, p. 56). For example, when undergraduate students start at university, they
may enter multiple social situations to which they may be largely unaccustomed. That student may become a housemate, a reveller, a sportsman, a learner, a society member, and so on. In such cases, the student learns these roles from individuals established in those cultures. Consequently, practices are transferred between individuals doing a similar activity.

There are times when these identities cause contradictions, particularly in relation to dividing time between activities. For example, the author recalls an instance when a student intended to miss an assessment in order to participate in a sports event. In this instance, the student’s identity as a sportsman contradicted the student’s identity as a student engineer; the sportsman was the dominant identity.

Colloquially, one might refer to wearing hats to represent such shifts in identity: an academic that is both a researcher and a lecturer might respond to a situation differently, depending whether she ‘puts her researcher’s hat on’ or ‘puts her lecturer’s hat on’. Likewise, a student that confronts a contradiction between his need to sit a test and his need to participate in a rugby competition may — consciously or unconsciously — choose which hat to wear.

Black et al. (2009) proposed the notion of leading identity to build upon Leont’ev’s (1981) earlier concept of leading activity. Leont’ev described leading activity as those activities that are more significant for an individual’s development. In other words, “activities become leading when new motives are generated so that the original motive of actions is surpassed by a new motive, and hence, a new activity” (Black et al., 2009, p. 57). They also argued that participating in a leading activity leads to a hierarchical structuring of existing identities according to their importance to the leading activity; they define ‘leading identity’ to be the identity that develops and dominates in the course of engaging in leading activity. They give the following example:

“...a student may tell of a particular work experience placement (e.g. in a hospital) as being critical in shaping their aspirations (to become a doctor) — this can be classed as a leading activity for the student if (a) the placement was begun with a particular motive (e.g. a week-long break from the toils of school) which then shifted to a new motive (to understand and engage with the occupation in question for its own sake) and (b) this resulted in a new leading motive for the student drive by one’s potential identity within that occupation. Within this framework, the student acquires a new leading identity (e.g. a medic) which is revealed by the new motive...” (Black et al., 2009, p. 58)
In this instance, the student’s new motive directs how his or her identities are ordered in this hierarchy of identities, prompting a new evaluation of his or her current and future activities — “e.g. what qualifications he or she needs to attain, what subjects are relevant” (Black et al., 2009, p. 58).

In the United Kingdom, most students choose which university courses they are interested in studying while in their final year of school and before they receive their exam results. Their motives could be aimed towards entering a particular profession, or pursuing the topic of their interests further, or to fulfil the wishes of their families, for example. During the course of their university studies, new leading motives may emerge: for example, this may happen as students learn which modules most interest and engage them. For these students, their university studies become leading activities and they acquire a new leading identity.

Some students may have already developed a leading identity during their school education that continues to have a noticeable effect on their identity formations at university. Those students may have chosen their favourite subjects to study at school to develop their interests in those subjects; however, the desire to score high marks may become the new leading motive, given that they may be perceived to satisfy the need of success. In such cases, the students may develop a new identity that prizes success in excellent performance in assessments and examinations. This identity may transfer into university study.

Leading identity directs the course of leading activity once it emerges. It also has an impact on the goals that the individual pursues: a student whose leading identity is to become an engineer may have different goals from a student who wishes to earn a first class degree, regardless of the course subject.

Thus identity is an important consideration when discussing the effectiveness of an assessment. It forms an important part of the activity of university learning and may have an influence on the forming of a leading identity. It creates a double bind of priorities when considering how assessment should cater for the student: should assessment cater for every identity, or should assessment be designed to nurture the development of a particular identity?

In other words, is it the responsibility to design assessment so that it suits the student, or should assessment be designed to train students to develop in a particular way? If assessment should cater for every leading identity, then it must cater for a wide range of possible goals. That is, as well as providing guidance to students wishing to seek a deeper understanding of the material being tested, an assessment should also give feedback that helps students aiming towards less demanding goals.
If assessment should be designed to nurture particular identities, then, as the subsequent discussion of motives and goals argues, there remains a risk that those students that do not develop those identities fail to engage and flourish with the assessment.

Cultural-historical activity theory accepts that individuals within a community have different identities, but in order for the individual to succeed in the community, they must act in accordance with the motives of the activity. Therefore, some identities and actions are valued more than others, and the community encourages individuals to orient their actions towards those of the group. Thus the community has an influence over the goals that individuals pursue when they are engaging in a common activity.

In order to understand how goals relate to the effectiveness of assessment, the following section discusses how goals are formed and what happens when goals are achieved or missed in the course of learning.

### 3.2.2 Goals

The literature review established the importance of the assessor’s goals to assessment in terms of setting assessment criteria and designing assessment. However, less is mentioned of the student’s goals in relation to assessment.

Perhaps this is because there is an expectation that the teacher or assessor is responsible for choosing the goals for assessment and for students to adopt these goals. For example, in an examination, the assessor will select items of the syllabus to be tested and design the test paper accordingly; students are bound to the goals of the paper in order to succeed.

Yet, students are individual actors with agency and with their own motives. They consequently form their own goals and act accordingly (Leont’ev, 1977). If the goals set by the assessor for an assessment do not align with the student’s goals for assessment, then there is a contradiction; in which case, the student may not engage fully with the assessment. An example where this kind of conflict between teachers’ and students’ goals can be found in the study conducted by Jaworski, Robinson, Matthews, and Croft (2012), where students engaged in learning activities that were set by the lecturer, but failed to acknowledge the benefit these activities had for their own aims to do well in summative assessments.

This raises some questions about the nature and origin of the assessor’s goals, when formulating an assessment, and the student’s goals, when engaging with an assessment:

- how are these goals formed and selected?
• how individual are the choices for goal-formation?

• what happens at the end of an action directed towards the achievement of a goal?

There is little in theoretical literature about the process of forming goals. Tikhomirov (1988, p. 113) noted, “The category of goal is constantly used in the theory of will, but the very process of formation of the final goal, described as decision-making, is at best mentioned but not treated as a subject for experimental research... It is also necessary to remember that the problems of will are among the least developed”. A search for a development or contrast of Tikhomirov’s ideas in existing literature yielded little.

However, while Tikhomirov discussed “thinking and goal-formation” within the context of Marxist and socio-cultural philosophy, he did not explicitly state that his notions referred to activity theory. Nonetheless, his terminology was consistent with activity theory literature and has been cited in the work of Engeström, and the notions he presented are supported in the literature of researchers using activity theory in their work.

For the theoretical framework of this research, Tikhomirov’s position is taken to define and describe goal-formulation and to answer the questions above.

**How are goals formed and selected?**

When individuals form new goals, there are a number of prerequisites (Tikhomirov, 1988). These include: “the emergence... of new needs and motives, assimilation of new knowledge about possible goals..., acceptance of new requirements for an action, appearance of new results of individual actions, failure to attain the anticipated results, emergence of new unconscious anticipations of future results of an action” (Tikhomirov, 1988, p. 117). This suggests there are many causes of goal-formation, which arise from the outcomes of an action or from increased awareness of the consequences of a future action. With respect to CAA, Table 3.1 gives some examples of scenarios that invoke goal formation.

Goals are also viewed as an anticipated result of an action. When considering possible future outcomes, the individual also has some sense of the likelihood of the desired result being achieved. Tikhomirov (1988) suggested that individuals may have numerous motives, and the goals they formulate reflect this fact. Sometimes, the motives offer clashes; in which case, the goal might not contribute to some aspect of the activity of some motives.
Chapter 3  Theoretical Framework

Prerequisite CAA Example

<table>
<thead>
<tr>
<th>Prerequisite</th>
<th>CAA Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>emergence of new needs and motives</td>
<td>student identifies need to get higher module marks in order to achieve desired degree classification through CAA, the student learns a particular technique that can be used to prove a theorem</td>
</tr>
<tr>
<td>assimilation of new knowledge about possible goals</td>
<td>student realises that in order to perform well in the exam, he/she must master a technique that is tested in CAA</td>
</tr>
<tr>
<td>acceptance of new requirements for an action</td>
<td>having established that he/she could do a particular task through a CAA test, he/she aims to complete a related coursework task using the same technique</td>
</tr>
<tr>
<td>appearance of new results of individual actions</td>
<td>a student aiming for full marks may concentrate on the questions he/she got wrong first time the student might anticipate that getting 100% in a CAA test would make doing a related coursework easier</td>
</tr>
<tr>
<td>failure to attain the anticipated results</td>
<td></td>
</tr>
<tr>
<td>emergence of new unconscious anticipations of future results of an action</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Examples of prerequisites that invoke goal formation in CAA

For students embarking on a course of study at university, many motives influence their goal-formation and selection. As discussed before, on the notion of “leading identity”, a dominant aspect of a person’s identity can be challenged by strong motives (Black et al., 2009). Black and Williams (2013) followed that paper with a description of the tensions that Mary, a Muslim woman that had ambitions to become an engineer, encountered. Once at university, she found her motive to get “to the top of engineering” (Black & Williams, 2013) was impeded by her identity as a “good Muslim woman”. Consequently, her goals were formed and selected to do well enough in her course without moving too far from her family home. This meant that she felt she could no longer reach “the top”, but she could remain close to her family.

In the context of this study, students bring leading identities that influence the way they act and the goals they pursue. Tensions, like those described by Mary, are likely to manifest as a result of motives and identities competing. For example, a student may face a choice between spending time to understand the material or to attempt to maximise their marks when doing a CAA test.

Tikhomirov (1988, p. 120) mentioned other ways in which individuals form goals; Table 3.2 gives examples of how such goals may be formed with respect to CAA:

(a) “transformation of motives into goal motives in the process of gaining an awareness of them” — when individuals become aware of their motives (deciding to become a pilot, for example), then they form their goals accordingly;
Chapter 3  Theoretical Framework

Ways of forming goals

| Transformation of motives into goal motives in the process of gaining an awareness of them | A student, during his/her degree course, may become aware of his/her career ambitions and uses these ambitions as goals. |
| Transformation of side-effects of action into a goal through a connection with the motive and gaining awareness of the result | Through CAA, a student might notice an alternative method for solving a problem; the student may set a new goal aimed at learning, testing and refining this method |
| Transformation of unrealised results into realised | Once a goal has been achieved, the student can formulate a new goal |
| Establishment of intermediate goals | The student might choose to concentrate on correcting wrong answers in order to achieve the primary goal of achieving a high score |

Table 3.2: Further examples of prerequisites of goal formation in CAA

(b) “transformation of side-effects of action into a goal through a connection with the motive and gaining awareness of the result” — in pursuing a goal, an individual may notice a related task that would be beneficial to perform;

(c) “transformation of unrealised results into realised” — once an individual has achieved a goal, another goal may form;

(d) “establishment of intermediate goals” — an individual may identify a long term goal that may require intermediate tasks to be completed.

How individual are the choices for goal-formation?

Individuals determine which goals they pursue, whether they formulate their goals themselves or accept socially generated goals. Of social and cultural influences on goals, Tikhomirov (1988, p. 117) noted: “An individual views a group goal as a requirement which has yet to be accepted. Moral and legal norms are a source of goals but should not be identified with them (since otherwise law-breaking could not have been explained)”. Whether individuals accept a group goal is determined by “selectivity, which is determined, among other things, by personal experience” (Tikhomirov, 1988, p. 117).

Although individuals may accept group goals for their own activity, individuals may also formulate their own goals that have some influence from social and historical factors. These factors include the needs determined by the social structure, social conditions and the mastering of language. Mastering language enables the individual to negotiate goal-formation both with the social group and internally. The individual, by means of these negotiations, can establish whether the conditions form a link with a current need, with
a potential need, or as a means to resolve a conflict between two needs (Tikhomirov, 1988, p. 118).

When students are presented with a learning activity, such as CAA, there are inherent goals that are presented by the lecturer. In particular, the lecturer wishes the students to attempt the assessment and improve. Furthermore, the lecturer will have some specific goals in mind when designing the assessment. There are also goals that are presented by the academic community: to test the students’ competences. When selecting their own goals, the students may have some awareness of lecturer and academic goals, while also being complicit in the generation of group goals among the student community — for example, if members of the group are finding a particular method difficult to grasp, the group might elect to practise those questions more.

One can conclude from Tikhomirov’s writing that individuals within a group form and pursue a somewhat unique repertoire of goals. In terms of learning, and with respect to using CAA, one would expect students set a personalised combination of goals, which results in contrasting activity. These goals may yet be quite similar, given their contextual nature: that is, they have been formulated within a given community and culture for a particular activity. However, since individuals acquire and develop a leading identity, one can expect differences in the goals that students select.

**What happens at the end of an action?**

Tikhomirov (1988) believed that there are at least three different types of results of an action. One is to achieve the goal. Another type of result is an unintended outcome of the action, which may arise from a successful or unsuccessful action. The other type is the outcome of an involuntary activity that is associated with the purposeful action. For example, a student answering questions on a CAA test with the goal to understand an aspect of calculus might or might not be successful in doing so (the first type of result); the student might learn other aspects of calculus in doing the test (the second type); and the student might learn a new keyboard shortcut for the CAA system (the third type).

Each of these three types can facilitate the formulation of a new goal. For the latter two types, a new goal may be formulated with respect to the unexpected or unintended outcomes, regardless of whether the original goal had been fulfilled.

This encourages the notion that selecting goals and performing actions form a cycle (fig. 3.3): that is, activity can be viewed as both a process and a cycle of actions. Tikhomirov (1988, p. 120) described this cycle in terms of the product of action: “an action is, as it were, doubly productive: the creation of a new material product is preceded by the
creation of a new ideal product.” By this, the author believes that Tikhomirov suggested that individuals do not achieve a sense of accomplishment and closure upon completing an action; rather, an opportunity for improvement immediately presents itself.

Tikhomirov (1988, p. 119) stated, “Viewed as a process, an action may be said to have three component parts: the formation of a goal, a progress evaluation, and evaluation of the final result”. That is, an action requires an initial goal, an assessment of whether that goal has been achieved, and reflection upon the action and the goal.

The individual reflects upon previous goals when performing a self-evaluation on his/her action. Tikhomirov (1988, p. 120) noted: “When a sequence of goals is formed, each is compared against previous results and previous goals.” Furthermore, Tikhomirov believed that the senses of success and failure are necessarily linked to this kind of introspection, since the individual has a sense of what he/she is capable of achieving.

This further alludes to the cyclic nature of activity and, therefore, learning. Not only does the completion of one action inspire the formulation of another, there is also a historicity to goal formation. When forming goals for a learning activity, students would be aware of their prior achievements, successes and failures when engaging in this type of activity. Perhaps a consequence of this suggestion is that the students are capable of setting goals at least as challenging as the ones they have achieved before.

Tikhomirov (1988, p. 146) added:

“Assessments of the result’s attainability, of which the person is not always aware, are incorporated into the process of forming a new goal. Apart from a conscious image of the future result, a real goal also includes an unconscious
assessment of the attainability of that result. These ideas have been confirmed by a series of experiments.”

In this quote, Tikhomirov suggested that students assess how realistic their goals are when they form them, yet they may not be aware they have made such foresights. Likewise, for CAA, a student would form goals and assess their viability concurrently. An outcome of such deliberations — conscious or unconscious — is that students will more likely pursue goals that they believe are achievable than they are to pursue goals that they do not believe are achievable.

Tikhomirov’s (1988) descriptions of the goal-forming and goal-achieving processes make five clear points that are important to consider when analysing the activity of students with an assessment tool.

1. A myriad of external factors shape how each student forms their own individual goals.

2. While adjusting to a new community and culture at university, students are exposed to new influences and group goals. That is, they are influenced by lecturers and peers in the goal-making process.

3. In order for a student to know whether a learning action has been successful, there must be an assessment to determine whether a goal has been achieved. Usually this assessment is designed and administered by a lecturer or assessor, but assessments may be peer-led or self-directed.

4. Once a goal is achieved, students are motivated to pursue more challenging goals. These iterations of improvements suggest that learning occurs in cycles.

5. Students use their histories, as well as their identities, cultures and social influences, to support their decision-making when forming and selecting goals. They assess whether goals are attainable while they are forming their goals. Most of the time, students set goals they believe they can achieve.

Tikhomirov’s writing suggested that the community of the activity has an influence on individuals’ goal formations. Understanding the nature and extent of this influence should be of interest in the design of assessment: if students change their motive in the process of doing an activity, it becomes a leading activity and the students develop leading identities. If this is the case, it suggests that students could be encouraged to develop new leading identities — that is, with motives that align more with the community activity — through exposure to and immersion with the community.
The following section explores the role of community on activity and how it may affect how learning and assessment take course.

### 3.2.3 The influence of the community

Engeström’s (1987a) triangular model of activity (Figure 3.2) makes clear that each node has a reciprocal influence and importance with respect to all the others. Community is one such node; thus the community in which an activity takes place not only affects the subject, object, tools, division of labour and rules of the activity, it has a telling influence on the process of the activity as well. Figure 3.4 offers an example of the activity of a student undertaking a CAA test.

This section on ‘community’ begins with a brief review of the impact that community has on the other nodes of activity, as described by Leont’ev, Engeström and others that have assisted in the development of understanding activity.

It continues and culminates with the notion of figured worlds, as proposed by Holland et al. (1998b), which offers a construct for framing and understanding how individuals act within particular socially and culturally formed contexts.

**Community in the activity system**

Leont’ev (1977) affirmed the critical importance of community in human activity:

“Despite all its diversity, all its special features the activity of the human individual is a system that obeys the system of relations of society. Outside these relations human activity does not exist.” (Leont’ev, 1977, p. 7)
While individuals are intrinsically influenced by the society and culture in which they are immersed, Leont’ev was clear that it was not a simple matter of deducing or predicting an individual’s actions by the norms of those societies and cultures.

In the previous section, *Motives and goals*, Tikhomirov’s position on the formulation of goals and the choices that individuals and groups make to pursue those goals were described. He believed that all group goals were once formulated by individuals, and individuals within those groups may elect to adopt or reject those group goals in their own actions. Pointedly, he noted that it was not simply a matter of individuals adopting the goals of the group — or community — to which they belong, “since otherwise law-breaking could not have been explained” (Tikhomirov, 1988, p. 117).

From Leont’ev and Tikhomirov, there is a clear impression that the individual and their community are not always reflected faithfully from one to the other. However, Leont’ev noted the importance of community and the adoption of a common motive across the community as a necessary characteristic of human activity derived from the Marxist ideology from which activity theory was conceived.

Engeström (1987a) made the influences of community on the activity system clear in his triangular representation of activity. In this diagram, community is seen to have a reflexive influence on the subject, the object, the division of labour, the rules and the development of tools.

He also suggested that “each sub-triangle... is potentially an activity of its own” (Engeström, 1987a) and used Marx’s (1999) terminology for *production, exchange, consumption* and *distribution* to represent them (Figure 3.5). The influence of community on the other nodes of activity is most telling in the latter three triangles.

![Activity in Marx (1999) as parts of the Engeström (1987a) triangle](image)
Marx (1999) described the four processes represented by the sub-triangles in the following way:

“...in production the members of society appropriate (create, shape) the products of nature in accord with human needs; distribution determines the proportion in which the individual shares in the product; exchange delivers the particular products into which the individual desires to convert the portion which distribution has assigned to him; and finally, in consumption, the products become objects of gratification, of individual appropriation. Production creates the objects which correspond to the given needs; distribution divides them up according to social laws; exchange further parcels out the already divided shares in accord with individual needs; and finally, in consumption, the product steps outside this social movement and becomes a direct object and servant of individual need, and satisfies it in being consumed.” (Marx, 1999, ch. 1, sect. 2, para. 2)

In this passage, Marx noted the community influence in all four processes. In production, the community and the individual agree on the object of the activity; in distribution, the responsibilities for individuals within the activity are allocated; in exchange — or, as Engeström suggested, communication — products, tools, rules and information are exchanged between individuals in the community of the activity; and in consumption, the product of the activity serves to satisfy the needs of the individuals within the community.

Hence it is clear that the community of an activity has a direct impact on the processes that occur within that activity. Consequently, the course and end-product of the activity are also affected by the community.

How does one define the community in which lecturers and students engage with computer-aided assessment? If the community is defined by those agents that have some investment in the individual’s activity — that is, those agents that have an effect on the individual’s activity — then the scope for those agents included is very large: students are influenced by family, friends, peers, lecturers, former teachers, and so on; for lecturers, colleagues, administrators, students, research literature, and so on, may have an influence on their teaching.

However, it is perhaps somewhat peculiar to consider the student’s family, for example, as part of the community in the activity of learning at university or taking computer-aided assessments. Nonetheless, there remain cultural and historical influences on the individual, regardless whether these influencers may be considered as part of the ‘community’ of the activity.
In order to make more sense of the distinctions between those in the community of the activity and those that merely form an external cultural or historical influence, the notion of *figured worlds* is useful as a framework for describing the agency of individuals both to be an individual within a community seeking a common motive, and to have multiple identities for engaging with multiple activities and multiple communities.

**Figured worlds**

Individuals are engaged in several activities in which there are different communities, expectations, outcomes and tools. In participating in each activity, individuals are expected to act according to those norms and accepted practices: in other words, individuals act differently according to the activity system in which they are engaged in. Holland, Lachicotte Jr, Skinner, and Cain (1998a) called this notion “figured worlds”.

“By ‘figured world’, then, we mean a socially and culturally constructed realm of interpretation in which particular characters and actors are recognised, significance is assigned to certain acts, and particular outcomes are valued over others.” (Holland et al., 1998a, p. 52)

A university comprises several figured worlds: some value academic endeavours; some reward sporting achievement; some desire administrative efficiency; some value social skills. Students and lecturers at the university find themselves engaging with several different figured worlds with activities that sometimes compete for an individual’s time and effort.

Lecturers are teachers, assessors and researchers. While teaching, they are governed by rules of the institution and expectations from the student; they engage primarily with students and use a variety of tools to communicate. While assessing, lecturers abide by a different set of rules and cultural expectations; they have a different role in administering and reporting students’ performances and use different tools. While researching, the community is very different — little interaction with students and more with peers and colleagues — and the motives are substantially different.

For students, university life transcends study. There are several study activities: in the lecture, in the seminar, in the library, in the exam hall, and so on. The activity is visibly different in each. Many students also live at university and, for some, their social lives are centred around their halls of residence. They might also engage with sports groups, clubs and societies.
Figured worlds are spaces in which new identities not only develop but are expected to develop: in order to engage with others and achieve while in a figured world, individuals must participate in community activities. However, the identities that develop within figured worlds tend not to be distinct or “brand new”. Consequently, one might expect that new students entering university will inherit aspects of their identities developed prior to arrival and develop new characteristics as they partake in these new activities.

Citing Holland et al. (1998a), Rush and Fecho described figured worlds as being continuous and developing; to which the same could be said for individuals’ identities as they develop in a figured world:

“Figured worlds acknowledge the simultaneity of their historic and the dynamic. As such, there is an understanding that what is will continue, but what is will also be different even as it retains qualities of what has gone before. The process of improvisation — of finding space for oneself within a larger social construct — figures significantly into this transaction of the historic and the dynamic.” (Rush & Fecho, 2008, p. 126, emphasis added for clarity)

Rubin (2007, p. 245) concluded: “The conceptual frame of figured worlds is a powerful tool for illuminating how everyday activities and events become part of identity production.” In the Rush and Fecho (2008) study, the students were encouraged to develop identities as speech and language practitioners; Rubin reported on the effects of a figured world on the development of students’ identities in the classroom.

The idea that figured worlds develop participants’ identities is an important one. The Rubin study indicated how this can have a negative impact on students’ identities and, as a consequence, how they form their aims, motives and ambitions.

The figured worlds from which university students come from typically — but not always — relate to secondary school, in which there are expectations for conduct, study and assessment. Upon arriving at university, these students arrive at new figured worlds in which the rules are not enforced in the same way, and the division of labour is such that the students are expected to work more autonomously.

As well as being a place for learning, university is also a social space. This study is situated in a British university with a large campus: many students live on-site or nearby. It is renowned for its reputation in sports and many students participate in societies and clubs. Each of these social spaces forms a figured world and students may form multiple identities.
Thus a student might not just be a student while at university: they might also be footballers or hockey players; they might be activists, fundraisers or student representatives; they might also be a halls resident and form their own communities based on their interests with housemates.

As these identities develop, students may find that their figured worlds collide: they may be faced with tensions that arise from sports fixtures or outings, for example. In such cases, the students may have to align more with one identity than another.

Lecturers face similar tensions and choices. They engage with many figured worlds within the university, which include research, teaching and assessment commitments. In each of these figured worlds, with differing — and sometimes conflicting — demands and outcomes, lecturers juggle their time to fulfil their commitments.

Therefore, an important consideration for choosing a methodology is to capture these figured worlds and how they impact on identity development and the resolution of tensions between competing identities.

It is possible that CAA could be considered a figured world in which individuals develop new identities and actions. If CAA does form a figured world, then there may be tensions that arise between the expectations within the CAA figured world and other figured worlds.

In order for CAA to be considered a figured world there must be identities, actions and goals that are valued distinctly from other figured worlds in the university community. Therefore, the analysis should determine what identities, actions and goals are valued in CAA testing and whether these are considered distinct and different.

### 3.2.4 Resolving contradictions

Over time, activity systems develop norms and customs. There become “ways of doing things”, accepted practices and expectations. However, there can be deviations from these norms and they are called “disturbances” (Engeström, 2000, p. 964). Disturbances are caused by contradictions (Engeström, 2001). Engeström defined contradictions in the following way.

> “Contradictions are not the same as problems or conflicts. Contradictions are historically accumulating structural tensions within and between activity systems. ...Such contradictions generate disturbances and conflicts, but also innovative attempts to change the activity.” (Engeström, 2001, p. 137)
Engeström (1987a) described four levels of contradiction. At the primary level, contradictions exist within nodes of the activity system. Foot (2014) argued that primary contradictions are the foundations of all contradictions and that they arise from the dichotomy of use value and exchange value. This is most easily conceived as conflicts in the object of the activity: for example, a student chooses a university course to pursue their interest in a field, but at the same time their degree holds significant societal value.

This dual purpose does not always present a problem, but it is an instability. Consequently, this primary contradiction provides a basis from which secondary, tertiary and quaternary contradictions might develop.

A secondary contradiction occurs between nodes of an activity. For example, a student might find a way of obtaining a solution to an assessment question using mathematical software. This is a contradiction between the tools (the assessment and the software) and the rules (plagiarism and academic honesty) of the activity system. The tension arises from the primary contradiction: there is a reward for obtaining the correct answer (exchange value); but there is also one for acquiring knowledge that would prove useful in a future career (use value).

Foot (2014, p. 340) noted: “It is important to note that secondary contradictions exist a priori to and independently of tertiary contradictions.” Tertiary contradictions arise as a result of the development of a more “culturally advanced” (ibid.) activity. Thus an effort to remediate the secondary contradiction between the software and the rules might include invigilating tests and withdrawing access to the software. This may cause tensions within the student body: particularly if they felt that an invigilated test increased pressure and impaired their performance.

Quaternary contradictions arise from neighbouring activity systems remediating a tertiary contradiction. Regarding the student studying on a degree course as the central activity, consider the activity of the lecturers: in assessment, they determine and report on the competences of their students. The students realise that they must achieve well in these invigilated tests to earn the credit they need; and therefore they change their study habits accordingly. The lecturers may interpret these new study habits as surface learning, lacking a depth in understanding.

Therefore, while contradictions offer an insight into past and future change, the level of contradiction is important for establishing possible causes of change.

Engeström (1987a) argued that contradictions and their resolutions provide the basis for learning. Engeström (1999) called this process “expansive learning” and described it as a cyclic process comprising seven stages: questioning, analysing, modelling, examining model, implementing model, evaluating process, and consolidating new practice. Foot
Levels of contradiction | Characteristics of contradiction levels | Corresponding learning action(s)
--- | --- | ---
Primary | Occurs between the use value and exchange value of any corner of an activity system | Questioning
Secondary | Develops between two corners of an activity system | Analysing, Modelling
Tertiary | Arises when the object of a more developed activity is introduced into the central activity system | Examining model, Implementing model, Evaluating process
Quaternary | Occurs between central activity and neighbouring activities, triggered by tertiary contradiction | Consolidating new practice, Questioning

Table 3.3: Levels of contradictions and corresponding learning actions, from Foot (2014, p. 342) and Groleau (2011) related the four levels of contradiction to each stage of the expansive learning cycle (Table 3.3).

Interpreting contradictions as a basis for learning in this way reinforces the notion that learning is cyclic and depends on the need to satisfy and create goals for learning. The primary contradiction, occurring within the individual’s mind, creates an unstable state. To resolve this contradiction, the individual creates a goal to be achieved in order to solve this internal questioning. Once a learning goal is achieved, the individual is faced with a quaternary contradiction with a neighbouring activity, thereby exposing the individual to further potential learning goals.

### 3.2.5 Summary

This section, looking at some aspects of cultural historical activity theory in more detail, provided some possible solutions to these questions listed earlier in the chapter.

- Individuals arrive at university with prior experiences, exposure to different communities: how does this form who they are and how they might act? (*Developing an identity*, p. 49)
- People engage in actions to pursue goals: how are they formed (*How are goals formed?*, p. 54), and what happens when they are achieved? (*What happens at the end of an action?*, p. 57)
- What effect does the community have on the individual in an activity system? (*The influence of community*, p. 60)
• What happens when there is a contradiction in the course of their activity? (Resolving contradictions, p. 65)

From the literature on activity theory, it would appear almost certain that each individual’s activity is unique, because they have developed a unique set of experiences and knowledge. These experiences will have shaped the identity of the individuals, and their understanding of their needs and their agency and motives would differ accordingly. Therefore, while individuals may appear to be engaged in the same activity, their actions may differ as they seek to reach a similar object.

The community has an important impact on the individual. It is the community that collectively dictates exchange value and reward; and use value is determined by the individual’s interpretation of the worth of completing an action, which has an impact on the community. This is notwithstanding the direct influences that the community has on the activity itself: from providing support by dividing labour, to dictating rules.

Contradictions are the drivers of change in the activity. All contradictions start from primary contradictions and may escalate. Contradictions need not be identified as “bad” or “hindrances”; they provide the motivation needed to drive opportunities for learning.

3.3 Proposing a model for effective assessment

The theoretical framework provides a focus for the study and informs the methodology. Viewing the use of computer-aided assessment at a university as a mediated activity performed within a community of learners and teachers not only provides hypotheses for what occurs during the activity, it also enables the researcher to select a methodology to capture such characteristics.

However, the literature relating to assessment and to cultural-historical activity theory identify similar patterns of learning that do not seem to have been linked explicitly in the literature before.

Black and Wiliam (2012) attempted to link formative assessment with cultural historical activity theory and noted:

“Most theoretical writings about pedagogy pay scant attention to assessment, formative or summative, and although some argue that classroom assessment is concerned with instruction rather than with pedagogy, this simply re-defines the problem of incorporation, but does not solve it.” (Black & Wiliam, 2012, p. 227)
Black and Wiliam had been responding to Perrenoud (1998), who felt that the former authors’ writing on formative assessment (Black & Wiliam, 1998) lacked a theoretical framework from which to make their assertions on the effects on learning. Perrenoud warned:

“Without a theoretical model of the mediations through which an interactive situation influences cognition, and in particular the learning process, we can observe thousands of situations without being able to draw any conclusions.” (Perrenoud, 1998, p. 95)

While Black and Wiliam (2012) considered these mediations in more detail, there remained features of Engeström’s (1987a) Expansive Learning model and wider assessment literature that were not considered. Furthermore, the focus was on formative assessment: a wider contemplation of assessment practice would identify why formative assessment permits students to develop while other practices might not.

A particular feature of formative assessment, which did not gain significance in the works of Black and Wiliam, is the opportunity for formative assessment to be part of a learning cycle. They claimed that formative assessment was effective at promoting pupils’ achievements (Black, 1998; Black & Wiliam, 1998), but these improvements were not related to existing knowledge and continuous development.

An important part of ensuring that students improve is feedback, and that students are able to interpret and use the feedback they are given. In earlier definitions of formative assessment given by Black and Wiliam, the onus is on the teacher to give useful feedback to the student. When this feedback is given to the student, the assessment is formative; whether the student acts upon this feedback is not relevant to this definition.

Returning to Engeström and Foot (Table 3.3), there are contradictions that develop as students identify gaps in their knowledge and understanding. They seek resolutions to these contradictions by engaging in tasks that allow them to question, model, test and respond to their previous learning. Resolving a contradiction and achieving a goal reveals a new gap in knowledge, much as Tikhomirov (1988) suggested for goal formation.

Feedback is a necessary part of this cycle. There is a summative aspect to feedback, inasmuch that it informs the student whether the aims for learning have been met — whether these aims have been set by the lecturer or the student. However, there is a formative aspect that informs the student of the way to next proceed (Sadler, 1989).

However, the definition for formative assessment offered by Black and Wiliam (2012) overlooked the need for students to demonstrate that they have learned from the feedback
they have obtained. From a CHAT perspective, assessment may be viewed as a tool for establishing whether a student has learned. It may be argued that feedback is wasted if students are not given the opportunity to, subsequently, demonstrate their learning.

Since assessment is a tool in the activity of learning, it has the purpose of mediating the journey towards a learning goal. This implies a *summative* function since it appears to be an assessment of learning, rather than assessment for learning with a formative purpose (Wiliam, 2011). Such assessment is construed to be the end of learning (Yorke, 2003), which is somewhat contrary to the cyclic nature of learning proposed previously.

Moreover, one of the criticisms of the divergent definitions for formative assessment is that it is not clear what should follow such an assessment. A simple learning cycle (Figure 3.3) suggests that, ideally, learners achieve their learning goals and move on to more challenging learning goals. This implies that the assessment must relate to the learning goals. Assessment must, therefore, give feedback in relation to the participants’ learning goals: it must inform the participant whether they achieved their learning goals or not (the *summative* function) and give cues for the participant to progress (the *formative* function). This model can be visualised in Figure 3.6.

A consequence of this model of effective assessment is that lecturers learn from assessment in a similar way. Black and Wiliam (2012) referred to teachers using what they have learned from an assessment to adapt their teaching. Similarly, lecturers learn about their students — and about the assessment tool itself — by assessing their students.

Indeed, lecturers and students operate in a community in the activity of learning and assessment. There is a division of labour throughout the learning process where roles change and develop over time. The student might need substantial input from the expert teacher to form appropriate learning goals in the first instance, but may develop sufficient autonomy and self-regulation skills in time to develop more individual learning goals. Similarly, the expert often designs assessment and feedback opportunities for students until they are sufficiently knowledgeable to be able to peer- and self-assess. The influences of the community and the division of labour are demonstrated in Figure 3.7.
This model is an idealised view of the learning process: clearly it is not always the case that assessment is appropriate for both the student and the lecturer and what they wish to achieve in the assessment. It is also apparent that the feedback given might be insufficient for the student to determine whether they have achieved their learning goals or not; and it might not be clear from the feedback what the “next steps” are.

Therefore, the model suggests that an assessment is still effective while the participant is able to progress along the paths in the diagram. Consequently, when an individual is unable to progress, the assessment is no longer effective. Assessment, like any activity, is an exchange between the subject and the object: regarding the activity as a success depends both on the subject’s agency and the nature of the object. Thus effectiveness is determined on an individual basis and success can neither be determined in absolute terms nor generalised.

The assessment and CHAT literature informed this ideal model for assessment, and describes when an assessment is effective and no longer effective. This raises the question of the scope of effectiveness for CAA: the model suggests that, given individuals have their own goals for assessment, each individual will have their own scope for effectiveness of CAA. However, there may remain discord between the student’s goals for assessment and what the community values. For example, while the student may be aiming to achieve a particular score in assessment, the lecturer may desire to gauge the student’s understanding of a particular concept — in which case, the student’s goal may be achieved while the lecturer’s goal is not.

This chapter explored cultural-historical activity theory as a theoretical framework for the use of CAA as a tool in the activity of assessing student knowledge. It suggested that learning is cyclic, but individuals have their own aims for learning, which develop from an individual’s identity, history, culture and their community. It culminated in a
model for describing the scope of effectiveness of an assessment tool with respect to both student and lecturer.

It informed the methodology, which must capture the identity, community, motives and goals, actions and responses for individuals that participate in computer-aided assessment. The following chapter describes how the methodology was designed to record data on these aspects of the activity of learning and how the model would be used to find the scope of effectiveness of CAA for the individuals using it.
Chapter 4

Methodology

The purpose of this chapter is to discuss the requirements of the methodology, imposed by the research questions and the theoretical framework, and the reasoned selection of appropriate methods. It concludes with a description of the techniques used for data collection and how the analysis was performed. The research questions for this research project are as follows.

RQ1. What makes an assessment "effective"?

RQ2. Under what conditions is the CAA system an effective assessment?

RQ3. What would make the CAA system more effective as an assessment tool?

The selection of an appropriate methodology for this project followed from the belief that the interpretative paradigm is most compatible with the theoretical framework. Activity theory holds that all activity is purposeful, as does the interpretative paradigm (Schwandt, 2000). Consequently, in order to understand the activity, the researcher must be in a position to understand the context and meaning behind the action.

Cultural-historical activity theory emphasises the roles that both the individual and the community have in activity: an individual subject acts within the context of that community (Engeström, 2001; Sannino, 2011). That is, by understanding the community context one can interpret the actions of an individual. Furthermore, by acquiring a collection of images of individuals’ actions, one can obtain insight of the collective activity.

The model of effective assessment proposed in Chapter 3 (p. 68) highlights the social nature of learning and the individual pathways of each subject in the activity of learning. Therefore, individuals have their own goals when performing any learning action; and their path towards reaching that learning goal, using the assessment tool to verify the
success of that learning, is also unique. Therefore, it is entirely reasonable to expect that there is no single optimal way to use computer-aided assessment. There may be certain characteristics of computer-aided assessment practice that suit some students rather than others, for example. In which case, it might be detrimental to some (yet helpful for others) to recommend a set of practices (Farnsworth, 2013).

Through this project, the methods seek to develop a greater appreciation of individual users’ motivations when using computer-aided assessment and how these motivations shape how they use CAA. This is in alignment with the interpretative research paradigm. The findings of this kind of research then ought to “serve as a paradigm or exemplar, illustrative of something much more general” (Ernest, 1997, p. 25). That is, the outcome of this project is a case study and an evaluation of the CAA system used at this particular institution.

4.1 A case study approach

Ernest (1997, p. 25) stated, “One of the special features of the interpretative research paradigm is its use of case study”. Merriam (2014, p. 40) defined a case study as “in-depth description and analysis of a bounded system”. That is, the unit of analysis defines the case study and it is from this unit of analysis that the methods are chosen.

Citing Yin (2013), Merriam (2014, p. 43) argued that “case study is a design particularly suited to situations in which it is impossible to separate the phenomenon’s variables from the context”. In the case of the individuals using computer-aided assessments, this is evident:

- The content of the CAA system’s question banks is unique to the institution;
- The theoretical framework states that the circumstances and histories in which an individual acts has an effect on an individual’s perception of their needs and they act accordingly;
- The model for effective assessment suggests that an individual’s goals are a unique interpretation of their need to succeed in the society and culture that surrounds them; thereby their approach to assessment is unique.

Thus there is no possibility of separating the students’ and lecturers’ activity from the sociocultural circumstances in which they are found. A case study approach allows for the reporting of these circumstances alongside the actions that the individuals have made.
There are strengths and limitations to using a case study approach. Case study provides a useful starting position for new research, examining social units in context — for all their complexity and the number of variables — and reporting these complexities in their entirety. They describe contextualised situations in great detail and do not seek to represent a general case (Merriam, 1998).

However, the argument that case study is too specific to be of any great value overlooks the reasoning for using case study in the first instance. Merriam argued:

“...single cases, experiments and experiences of Galileo, Newton, Einstein, Bohr, Darwin, Marx and Freud, ...point that both human and natural sciences can be advanced by a single case.” (Merriam, 2014, p. 54)

Nonetheless, the effect of this limitation can be reduced by making several case studies. Similarities that exist between the case studies provide the basis for establishing patterns and hypotheses that may be tested with different methodologies in subsequent studies. Yin (2010) referred to this practice as “analytic generalisation”. Since case study provides a rich, holistic account of existing practice and does not involve testing hypotheses, it is useful as an approach to a context in which the problems may not be immediately apparent.

This was the case for evaluating the CAA system. There had been anecdotes shared between lecturers — and, possibly, between students — about the CAA system and how it is used. These informal anecdotes created some concern about how some individuals used the system, which gave rise to the need to evaluate how it is used. However, these anecdotes gave suggestions of the behaviour of a minority of students and little about the activity of the lecturers; there was no basis for generating hypotheses prior to the study. Therefore a case study approach created a point of entry to the evaluation.

Merriam (2014) noted that there had been criticisms of the reliability, validity and generalisability of case study research, due to the role of the researcher. The researcher is regarded as subjective and, therefore, potentially biased in the analysis and reporting of the data. However, to counter this argument, Merriam noted that case studies are less discriminatory in what is reported and they do not attempt to simplify the complexities of the context.

As for the researcher, Merriam made the following observation.

“Qualitative case studies are limited ... by the sensitivity and integrity of the investigator. The researcher is the primary instrument of data collection and
analysis. . . . But training in observation and interviewing, though necessary, is not readily available to aspiring case study researchers. Nor are there guidelines in constructing the final report. The investigator is left to rely on his or her own instincts and abilities throughout most of this research effort.” (Merriam, 2014, p. 52)

While it has been the experience of the author that training does exist in this methodology, such in-house training is not always readily available elsewhere. The author also argues that it is impractical for many research students to practise designing, conducting and analysing for case studies beyond the scope of their main research project. Therefore, the main research project constitutes the primary source of development and practice of this approach, though feedback is available through conferences, workshops and discussions with other researchers.

While Merriam argued that researchers are “left to [their] own instincts and abilities”, the presence of theoretical training and diligent use of methodology literature ought to provide adequate preparation for the use of a case study approach. Consequently, choosing and employing methods for a case study approach requires similarly careful reading and preparation.

### 4.2 Choosing methods

Choosing to perform a case study does not immediately yield a set of methods or procedures to be followed in order to conduct a study: “[it] is not a methodological choice but a choice of what is to be studied” (Stake, 2000, p. 435).

Merriam suggested that there are three main methods of data collection techniques in case study design: interviews, observations, and document data mining. The choice of methods from these options is informed by the literature, the theoretical framework and the research questions; but there are other considerations and constraints, such as time and cost.

Nonetheless, “in education if not in most applied fields, interviewing is the most common form of data collection in qualitative studies. In some studies, it is the only source of data” (Merriam, 2014, p. 86, emphasis in original).

The artefacts produced by CAA testing could be considered documents that could be mined for data. While these documents give some idea of the actions of the lecturers and students, the context of these actions and the details of these actions themselves are not recorded by the CAA system. It was possible to obtain data relating to the length
of time students spent on practice tests and the number of practice tests attempted, for example; likewise, data on the location and times that these tests were attempted were also available. On reflection, the author believes that there could have been some value in obtaining such data in relation to the case study participants. However, there were two concerns that precluded the pursuance of these data.

The main concern was that it was not a straightforward task to isolate individuals’ CAA activity from the system log files. Indeed, it transpired to be sufficiently difficult to obtain raw CAA summative test scores for cohorts to be discouraged from doing so. A request for individuals’ data would have taken a significant amount of an administrator’s time. There was also an ethical concern about obtaining an individual’s data — it would have had to be with the participant’s permission, and this might have had an effect on the recruitment of participants. Given that there transpired to be some difficulty with attracting participants, this was a risk that was not worth taking.

Having ruled out using artefacts from CAA testing, observations were also deemed to be unsuitable for this study.

4.2.1 Ruling out observations

Observations would appear most compatible with the theoretical framework. Kaptelinin and Nardi (2006, p. 31, emphasis added) stated:

“Activity theory maintains that no properties of the subject and the object exist before and beyond activities (e.g. Leont’ev 1978). These properties do not just manifest themselves in various circumstances; they truly exist only in activities when being enacted.”

Jonassen and Rohrer-Murphy (1999, p. 68) argued this case: “first and foremost, whatever the focus, the activity must be studied in real-life practice with researchers as active participants in the process”. Kuutti, Iacucci, and Iacucci further underline why active participation is important:

“One of the epistemological formulations in activity theory has been developed by V. V. Davydov, who separates empirical and substantial generalisations in the following way. Empirical generalisations are based on external observation and they are formulated by classifying and naming observable features into groups that are similar in some respect. The significance and relations between these classes are difficult to identify and uncertain, and the
generalisations themselves are static and lifeless. Substantial generalisations must be based instead on purposeful interactions with the material to be studied. Only in such interactions the hidden connections, dynamics and resistance will reveal themselves. Substantial generalisations are thus systemic and dynamic, and further actions can be based upon them.” (Kuutti et al., 2002, p. 97, emphases in original)

However, there are practical and ethical challenges to using observations in such a study where there are human participants undertaking a task that has long term implications on their success when they are either studying at or working for a university.

How students use computer-aided assessment is one focus of the project. For a simple observation to be possible, the researcher would need to know where and when students use computer-aided assessments. However, since all students are permitted to access the practice tests at times and locations of their choosing, this would require the researcher to be a participant in the social setting. From ethical and practical perspectives, this is not acceptable. Students may undertake these tests spontaneously, at any time of day or night, with others or on their own, and in locations not accessible to the researcher.

There is also a problem with regard to the reliability of the data collected. It is quite likely that a student would modify their actions for the presence of a researcher, particularly if they would normally engage in practices that would be considered plagiarism or in some way contrary to expectation. Observations conducted without the student being aware of the presence of the researcher may yield more reliable data, but would be ethically and practically unacceptable.

Similar arguments apply to the lecturers’ case as they prepare CAA tests for students or create new questions for the question bank. It is possible that lecturers might adjust their practices if they were aware of the presence of an observing researcher, for example.

Given that observations would likely be obtrusive and would manipulate the actions of both students and lecturers, they were not suitable methods for this study. Although some of the context and detail available in observations would be lost by using interviews to collect data, interviews were less obtrusive than observations and offered the opportunity to ask students and lecturers about their thoughts and past experiences with regard to CAA.

While interviews would give the level of detail that a case study would require, the author used questionnaires to gain a broad understanding of the society and culture in which individuals were subjects in the activity of learning at university. The questionnaires
were distributed to get an understanding of the population before interviewing a sample of students and lecturers.

Students are the most numerous end-users of computer-aided assessment. Most students will only ever see the front-end of the system in the test and practice areas. Yet, over the course of one academic year, students’ experiences differ. These are, in part, explained by the different practices they experience, but they also depend on their histories and the culture of the social groups they have become immersed in.

The questionnaires sought to provide some context for their histories and the activity while the interviews served to gain deeper insight into the perceptions and experiences of a small selection of individuals in that population.

For lecturers, computer-aided assessment seems to offer an ideal solution to a common problem. Students often need encouragement to work between lectures, and while assessment offers this encouragement it can often be time-consuming. Computer-aided assessment gives students both the motivation to obtain the marks that are on offer and feedback for the work they do. The feedback is an important consideration — it would be time-consuming to offer a similar feedback service if assessing “offline”.

This lecturer study considers how CAA might be effective for lecturers. It observes the emergence of different practices of students and the reasons for this divergence. A questionnaire on practices and opinions towards CAA was followed by interviews to explore the reasons for these practices and opinions.

### 4.3 Questionnaires

Questionnaires are used to collect structured survey information, largely using closed questioning to obtain easily classifiable data about the sample. The questionnaires in this project were used to develop awareness and understanding of how CAA was used. This awareness and understanding would then be used to inform the subsequent interview studies.

The questionnaires were designed following advice from Cohen, Manion, and Morrison (2013).

1. Decide the purposes/objectives of the questionnaire.
2. Decide the population and the sample (as questions about their characteristics will need to be included on the questionnaire under ‘personal details’).
3. Generate the topics/constructs/concepts/issues to be addressed and data required in order to meet the objectives of the research (this can be done from literature, or a pre-pilot, for example, focus groups and semi-structured interviews).

4. Decide on the kinds of measures/scales/questions/responses required.

5. Write the questionnaire items.

6. Check that each issue from (3) has been addressed, using several items for each issue.

7. Pilot the questionnaire and refine items as a consequence.

8. Administer the final questionnaire.

(Cohen et al., 2013, p. 379)

The population was determined to be the first year mathematics and engineering cohorts of students and the lecturers leading their modules. They were the primary users of the CAA system. The questionnaire was intended at this time to represent the population and, therefore, were distributed to every student and lecturer in this population.

Two questionnaires were designed to be distributed to the students: one at the start of the academic year; and one in the middle of the second semester. The intention was to identify any substantial changes in opinion and to capture experiences of using the CAA system during the course of the year.

As Cohen et al. (2013, p. 377) explained, “Respondents cannot be coerced into completing a questionnaire. They might be strongly encouraged, but the decision whether to become involved and when to withdraw from the research is entirely theirs.” The questionnaires to students were — aside from one cohort — distributed at the start of lectures, with one delivered online since lecture time was not available. In each case, the option to withdraw from the study or to opt out of the study was made clear on the questionnaire and in communication beforehand.

The questionnaires were used as a point of reference when designing the interview protocols. Therefore, closed questioning provided adequate data. The researcher’s supervisors provided feedback for both student questionnaires and were piloted with a second year group that were not part of the population considered for this project.

The questionnaires distributed to students (Appendix B), with 1,057 completed questionnaires received (688 from the first semester and 369 from the second semester). The data were analysed using SPSS. The method for analysis was largely the production of summary statistics and graphical representations of the data, with some comparison
statistics between cohorts and between semesters from the 285 students that completed both. While these figures did not contribute to the case study, the answers given by the interviewed students in the questionnaire informed the design of the interviews that followed.

The student questionnaire study highlighted differences between lecturers’ practices, supported by some anecdotal evidence in conversations with lecturers not involved in the study. Students in earlier focus groups also raised the issue that their lecturers’ policies on collaboration in CAA were not inherently clear (Broughton et al., 2011).

The objectives of the lecturer questionnaire were to gain lecturers’ professional judgments of computer-aided assessment and to obtain the practices currently being used in first-year mathematics and engineering modules at the target institution. Thus all thirteen lecturers of the mathematics and engineering mathematics first-year modules were approached to complete the questionnaire.

The questionnaire for lecturers covered several topics: the access that lecturers give students to the practice tests and the summative tests; lecturers’ policy on collaboration in CAA; what CAA tests of students; why lecturers use CAA; and what experience lecturers have had in developing the CAA system.

Since the intention was to use the questionnaires as a point from which to conduct interviews, it was sufficient to ask closed questions with a view to eliciting further information in the interviews. The questionnaire was presented to the researcher’s supervisors, who use the CAA system with second year engineering students and offered suggestions for improvement before the pilot. The pilot was conducted with PhD students to check for errors and to ensure that the questions were clear.

The questionnaire (Appendix C) was presented to the thirteen lecturers that were leading first year modules in mathematics and engineering. Three lecturers reported that they did not use computer-aided assessment. Nine questionnaires were returned completed. Their responses were used to inform the interview design.

### 4.4 Interviews

There is a precedence for using interviewing in higher education research, with cultural historical activity theory, and in mathematics education research.

Anthony, Haigh, and Kane (2011) used interviews alongside a cultural-historical activity theory framework in order to obtain data to understand the activity of induction for new
secondary school teachers in New Zealand. In this study, 100 newly qualified teachers participated in three interviews each, following a national survey.

The Transmaths projects (Black & Williams, 2013; Hernandez-Martinez, Williams, & Farnsworth, 2011b; Hernandez-Martinez, 2016), which explored students’ transitions to learning mathematics at university, used interviews as a means to collect data about students’ experiences. They used cultural historical activity theory as a framework for students’ activity and the notion of leading identity to model how students’ identities are formed during their transition to university.

Solomon and Croft (2015) interviewed 15 second-year mathematics students about disillusionment and disengagement with studying mathematics at undergraduate level. They subsequently analysed the transcripts to answer their research questions.

These recent studies made use of interviewing to develop narratives for each individual. These narratives delivered personal perspectives from individuals facing similar circumstances. From these interviews, the researchers identified those experiences that were common and those that were unique. The unique perspectives provided more information about the individual and how they acted under a particular set of circumstances; the common perspectives yielded information necessary for the case study.

Cohen et al. (2013) suggested that researchers must plan sufficiently prior to conducting an interview in order to avoid being the limiting factor for success. They identified seven stages of the interviewing process that must be considered, which also encompasses the considerations needed for the interviews themselves and the subsequent analysis. These seven stages — thematising, designing, interviewing, transcribing, analysing, verifying and reporting — are described here.

### 4.4.1 Thematising

The thematising stage of the interview planning process occurs at the beginning of a project. Deciding to conduct interviews — or focus groups — stems from the aims of the project, the theoretical framework and its perceived practical value. Interviews offered the students the chance to explain the reasons behind their choices in the questionnaires and to give a more detailed narrative of their experiences.

Interviews were chosen above other methods because they permit the researcher flexibility in the data collection process: though a structure exists, an interviewer can deviate from the scripted questions to acquire more detail. This is particularly useful as students begin to talk in terms of behaviours and routines.
The research questions seek the effectiveness of CAA in relation to students’ and lecturers’ goals. Therefore the interviews had to establish:

- what students aim to achieve;
- whether CAA helps them to achieve their goals;
- how they use CAA to achieve their goals.

While interviews were used in order to discuss the responses given in the questionnaire, they also provided the opportunity to ask some further questions. The initial aim of the interviews was to understand why students and lecturers used the CAA system in the manner that they did, and how their practices developed while using it. Having established the their practices, the interviews then sought to establish the extent to which the CAA system was effective for their aims.

4.4.2 Designing

In designing interviews, Cohen et al. (2013) encouraged naming the variables that are to be measured before designing the questions that will be put to the participants in order to measure them.

They suggested that there are three types of question-answer items: the fixed-alternative, the open-ended and the scale. In fixed-alternative items, participants are given a limited number of given responses from which they choose the most appropriate answer for their situation. This makes the coding of responses simpler; but respondents might not be so readily ‘pigeon-holed’, particularly if they feel there is an alternative to the options given that more suits them. As such, they are not so appropriate to engage participants in an interview.

Open-ended items give the participant the freedom to express thoughts in the manner of their choosing. However, it also gives licence for the respondent to reach multiple conclusions — as they are not compelled to choose from a range of responses — and this makes coding more difficult. To alleviate this problem, further questioning allowed the interviewer to probe further and clarify issues that the respondents raised.

Scale items present respondents with a range of answers they can choose, indicating agreement, rank, rating, and others. While this removes some of the coding problems that open-ended questions present, it also has similar problems to those presented by the fixed-alternatives. Therefore, scale items were not used in these interviews.
4.4.3 Interviewing

During the interviewing stage, it is important that the atmosphere is conducive to putting the respondents at ease in order to encourage the participants to give frank, open and honest responses. Cohen et al. cited Kvale (1996, p. 147) when explaining the importance of the interviewer for creating such an environment:

“Kvale adds that, as the researcher is the research instrument, the effective interviewer is not only knowledgeable about the subject matter but is also an expert in interaction and communication. The interviewer will need to establish an appropriate atmosphere such that the participant can feel secure to talk freely.” (Cohen et al., 2013, p. 422)

Cohen et al. (2013) continued to explain three elements that the interviewer should consider when preparing for interviews.

The first is the cognitive preparation: that is, the interviewer should be sufficiently familiar with the topic of discussion prior to attending the focus group or the interview. The second aspect is the ethical one. The respondents must be reassured that their identities and data are treated in a secure manner.

Thirdly, the interview dynamic must be considered. Although the interviewer has control of the interview, he or she must ensure that this control does not have an overbearing influence on the interview. Whyte (1982) devised a scale of the “directiveness” of the interviewer when responding to an interviewee, to which Kvale (1996, p. 135) added “silence”. There were occasions in the interviews where silence encouraged the participants to elaborate on or continue with their responses. Cohen et al. (2013) suggested that it is not necessary to feel bound to be directive or not directive over the course of the whole interview. Although abruptness may deter an interviewee, on some occasions it may be necessary to change the approach in order to get specific information or to help with later coding.

The students had earlier completed questionnaires on the subject of their experiences using CAA. These questionnaires offered some insight into students’ behaviours, insomuch that they confirmed that some students were making more use of tools and resources — such as textbooks, notes, the mathematics support centre — but little detail was obtained on why and how they used these resources. The questionnaires provided some background information that informed the design of the interview protocol and helped with classifying students for sampling.
What students aim to achieve | Whether CAA helps them to achieve their goals | How they use CAA to achieve their goals
--- | --- | ---
Students’ attitudes towards learning and university life, and other aspects they describe about their identity | The degree to which CAA supports the aims of the student and whether the student feels that CAA is effective | Students’ CAA routines, including any methods, supporting materials or techniques that help with CAA
Changes in students’ perceptions during the course of the academic year | Positive CAA experiences | Attitudes towards collaboration or instances of collaboration that students engage in

Table 4.1: Student interview themes identified from the questionnaires

The analysis of the questionnaires identified some key themes to explore in the interviews; Table 4.1 lists these themes under the three key questions posed in the Thematising section (p. 82).

Each student participant was presented with a unique set of questions. These questions were influenced by the responses given in the questionnaires. Each participant was asked variations of the standard questions plus others, depending on their questionnaire responses (Appendix D).

Emails were sent to the whole student cohort that took part in the questionnaire study to invite students to take part in interviews. The invite included an offer of a small monetary remuneration for inconvenience. Twenty students replied, from which ten were invited. The ten students represented an appropriate variety of mathematics and engineering students, and had all completed both questionnaires. Nine students were invited to participate in interviews; one of the invited candidates withdrew from the study.

The student interviews were conducted in one of the mathematics department buildings and in the university library. The participants were reassured both in writing and verbally, in person, that their data will be kept anonymously, confidentially and securely. The environment was designed to be a relaxed, social encounter. The first question helped to achieve this, but participants were also offered refreshments before starting the interview.

For the lecturers, the questions were based upon the questionnaire that they had recently completed. Each interview was tailored according to the questionnaire responses given. The standard questions are given in Appendix E.

The interviews were all recorded using a mobile telephone device for enhanced sound quality and portability.
4.4.4 Transcribing

The problem with transcribing an audio record of the interview is the data loss: by its nature it cannot capture non-verbal and visual communication between participants and the interviewer. Because of this loss, a transcription as a data source should be used with caution, as details given by non-verbal and visual means may be important in describing the context in which the verbal communication was made (Cohen et al., 2013).

In addition, transcribing needs to be faithful to the proceedings of the focus group or interview. Cohen et al. (2013) suggested that a moderator could summarise responses during an interview, but this too is a subjective view, and it is not always possible to capture non-verbal and visual communication. In any case, an audio recording captures much of the verbal communication.

In order to maximise the reliability and validity of the interview studies, several considerations were made in the design and analysis of the interviews in accordance with the advice given by Cohen et al. (2013).

- The participants were invited to take part in the studies at times of the year when they were least inconvenienced.
- The audio files were created on a mobile telephone device to obtain clear voice recordings. They were sent to a professional transcriber to transcribe.
- A single, professional transcriber produced transcripts for the interviews to minimise errors.
- The researcher coded the transcriptions consistently according to the research questions, the nodes of Engeström’s triangle, and to the learning cycle model of effective assessment.

The transcripts were coded in three ways:

- with respect to the nodes of Engeström’s activity triangle;
- with respect to the nodes of the learning cycle model for effective assessment;
- with respect to the research questions.

Each unit of speech in the transcripts that carried meaning relevant to the variables were coded. These codes, and their corresponding quotes, were used to develop profiles of the students, the lecturers, and their activity.
4.4.5 Analysing

Cohen et al. (2013) argued that, since the relationship between the interviewer and the research is unique, the analysis of the data is unique to the researcher:

“In qualitative data the data analysis here is almost inevitably interpretive, hence the data analysis is less a completely accurate representation (as in the numerical, positivist tradition) but more of a reflexive, reactive interaction between the researcher and the decontextualized data that are already interpretations of a social encounter.” (Cohen et al., 2013, p. 427)

Analysis was performed by coding units of speech that were deemed to carry meaning relating to the research questions, the nodes of Engeström’s triangle and the model for effective assessment. In the analysis, these codes were used to develop profiles of the individual students, constructing a narrative from the coded items. This narrative describes the students’ actions and use of CAA, describing student activity in terms of CHAT; the learning cycle model for effective assessment was then applied to these narratives to determine where the CAA had been effective for the student and the point at which the effectiveness ended.

4.4.6 Verifying

The verifying stage in the interviewing process requires examining the entire process for validity, reliability and generalisability.

In both quantitative and qualitative methods, it is impossible to achieve unequivocal validity. In quantitative approaches, one has to acknowledge standard error; in qualitative research, researchers and participants are necessarily subjective and so there will inevitably be some trace of bias. Instead, the researcher seeks to minimise invalidity and maximise validity.

However, the meaning of what it is to be valid is contested in Cohen et al. (2013). While some argue that the researcher, as the prominent and most immersive research tool in the process, can sufficiently validate the research, others argue that the subjectivity of the researcher makes interpretation fallible to bias. Interviews are, then, regarded as a co-construction between the interviewer and the participant (e.g. Mann, 2011). Thus further guarantees should be made to defend the validity of such interpretations.

Cohen et al. (2013) discussed how reliability can be achieved in qualitative research. They suggest the replication that quantitative methods strive for is not what is desired
in qualitative methods. Since a naturalistic approach observes a unique situation, one cannot expect observations to replicate. One might, however, expect that some sort of replication of an observation might be achieved by changing the researcher, or by focussing on different phenomena during the observation, or by changing the time or place of the observation.

In order to reduce bias and maximise validity, the interviews were prepared and conducted to follow the advice given by Cohen et al. (2013).

- A pilot interview with a lecturer of a second year module was used to test the effectiveness of the interview protocol and to obtain feedback for interview techniques.
- Students were invited to participate shortly after the Easter vacation and in good time before the summer examinations to ensure that students still had CAA in mind and that they would not be distracted by revision.
- The interviews were recorded on a mobile telephone device for enhanced sound quality.
- The audio files were submitted to the transcriber shortly after the interviews had been conducted. The interviewer verified the transcriptions.
- The transcriptions were coded in a consistent manner by the same interviewer.
- The participants were given the opportunity to provide further comments at the end of the interview.
- The researcher used the information acquired from the questionnaires to inform the design and delivery of the questions in the interviews.
- Although each interview was tailored according to the questionnaire responses, the structure of the questions and their order in the interviews were consistent.

Cohen et al. (2013) suggested that minimising bias is the most effective way of ensuring reliability in interview situations. They suggest that certain factors should be carefully considered so that bias performs a minimal role in influencing the outcomes of a focus group or interview. Leading questions can influence how a participant responds. There is also the perception of power; although either interviewer or participant may take power in the focus group, it is the responsibility of the interviewer to diminish the asymmetry of power. Structuring questions and building a rapport with the participants can help maintain a “power balance”.
Using Kvale’s (1996) advice for an effective interviewer, the researcher had developed thorough knowledge of the CAA system from the perspectives of the students, the lecturers, the developers and the administrators. During the interview, the researcher made notes to refer to when recollecting an earlier point to seek clarification or elaboration.

4.4.7 Reporting

The reporting stage should also reflect the study fairly. This requires the report to be transparent and detailed so that a reader is able to judge the fairness of the study.

The final stage of the planning process is the production of a report. In qualitative reporting, quoted items from the transcription should support points in the text; be accompanied with interpretation and explanation; be exemplar quotes that support the point made; indicate how they have been edited; and be incorporated within the natural flow and style of the report (Kvale, 1996).

For this report, there were specific questions that the interviews sought to answer about the students and lecturers in the study. The theoretical framework indicated the individuality of the participants and their actions in the study, thus each participant’s interview transcripts were analysed separately and reported separately. These separate analyses were used to develop collective analyses for the students (Appendix G, p. 183) and for the lecturers (Appendix H, p. 220), which emphasised the commonalities and differences in the participants’ responses. The research questions were answered using this combined analysis.

4.5 Summary

This chapter described the process for deciding the methodology and the methods used for collecting data.

The choice to use a case study methodology for this project followed from the selection of the theoretical framework, the research paradigm and the context of the study. The case study approach itself did not identify the methods of data collection that should be used; rather the use of interviews provided the opportunity to elicit details that would not be available from questionnaires, observations or data-mining.

Questionnaires were used as a means to understand the context of the use of the CAA system among the population, however. Students and lecturers were both approached to answer largely closed questioning to provide this information.
The sample of students and lecturers that participated in the interview study were not necessarily a representation of the population. However, interviewing several individuals enabled the emergence of common themes and disparate themes among the participants. These themes would enable common causes for the CAA system to cease becoming effective and to identify practices that allow the system to remain effective for longer.

The following chapter describes the analysis of the data collected from these methods.
Chapter 5

Analysis

The purpose of this chapter is to present the findings from the questionnaires and interviews conducted with students and lecturers. The nine students and six lecturers, from across the engineering and mathematics departments at the institution, formed a case study of the users of the computer-aided assessment (CAA) system. This case study answers the following research questions.

RQ1. What makes an assessment “effective”?

RQ2. Under what conditions is the CAA system an effective assessment?

RQ3. What would make the CAA system more effective as an assessment tool?

The analysis begins with the student study. As described in the Methodology chapter (Chapter 4), questionnaires were distributed to first year students in the seven engineering departments and the mathematics department. A brief analysis of the student questionnaire responses (Appendix F, p. 172) demonstrated that students had different experiences and views of the CAA tests, and had different aims.

Nine students were interviewed for this case study. Each of these interviews was analysed individually with respect to the research questions, with respect to the activity system, and with respect to the model for effective assessment. This chapter includes a synthesis of these analyses in order to answer the research questions.

The lecturer study included a questionnaire, to collect lecturers’ practices with respect to CAA, and interviews with six lecturers, to answer the research questions with respect to the lecturers. This chapter includes a synthesis of the individual interview analyses.
5.1 The student study

The overarching aim of this study was to ascertain how effective the computer-aided assessment (CAA) system used at the studied institution was for the students that were using the system. In the target institution, every first year student in the mathematics and engineering departments was expected to use the practice tests offered by the system in order to prepare for summative tests, which were conducted either online using the same system or on paper.

The cultural-historical activity theoretical (CHAT) approach to framing the development of knowledge and understanding, with the constructs that comprise the model for effective assessment, indicate the need to discover how the students approached the CAA tests and the influences on the actions that students take.

The Methodology chapter explained the importance of interviews in gaining such insights: the transcripts of several students’ interviews provided a detailed account of their activity and offered the opportunity to study similarities and contrasts between their approaches.

5.1.1 Student questionnaires

Students that were in their first year of mathematics and engineering degree studies at the studied institution were approached with two questionnaires. These questionnaires provided some background information about the students’ prior experiences of being assessed online (Appendix B.1, p. 160) and then their experiences of CAA at the studied institution at the end of their first year of study (Appendix B.2, p. 163).

In total, 285 students completed both questionnaires. Table 5.1 shows the number of students from each of the mathematics and engineering schools completed both questionnaires. Nine of the students that had completed both questionnaires were invited to take part in individual interviews. The recruitment of a sample that reflected the population was hampered by a low response to the invitation to participate in interviews, despite the offer of inconvenience payments. The distribution of interview participants is also described in Table 5.1.

A more detailed analysis of the student questionnaire study is given in Appendix F (p. 172). The questionnaires gave some insights into the views of the students towards the CAA tests.

- Students tended to find the CAA tests straightforward.
- The tests did not offer as much feedback as many students would have liked.
subject number eligible subject number participated

Engineering Department A 57 3
Engineering Department B 37 0
Engineering Department C 3 0
Engineering Department D 31 0
Engineering Department E 15 1
Engineering Department F 43 0
Engineering Department G 9 1
Mathematics 90 4

Table 5.1: A breakdown of the students that completed both questionnaires and those that took part in interviews.

- Many students used the practice tests to improve their test scores, and many used these practice tests until they had achieved the score they would like in the summative test.

- Some felt that coursework was a better test of their conceptual understanding than CAA testing.

- Students appeared to be less-inclined towards collaborative working on practice tests than they would in other assessments.

- Not all students aligned with these views and offered contrasting perspectives.

These findings offered an impression of the conditions of the CAA testing activity, and of the prevailing attitudes towards the CAA tests. The students appeared to be relatively satisfied with the CAA testing system, but felt that the system lacked the efficacy to offer the level of feedback they desired or the level of challenge they would expect.

However, examining individual students’ responses in each questionnaire gave a more detailed perspective of each student’s activity when performing CAA tests. These proved useful in the design of individualised interview questions: the interviews built upon the information acquired in the questionnaires to understand the students’ identities, aims and actions.

5.1.2 Student interviews

The analysis of the student interviews involved establishing the students’ identities, their aims and their actions, before using the students’ exposition of their practice to identify to what extent the CAA tests were effective assessments for them.
Figure 5.1: Alpha was largely satisfied that the CAA tests helped her to get the outcome she aimed for. However, she also felt that the system did not always help to identify her weaknesses.

For brevity, the thesis contains one case study from the student interviews to exemplify how the analysis was performed. The nine interviews identified commonalities and differences between students and their approaches to the CAA tests. They are presented first in sections describing the students’ identities, their aims and their actions. Using the case studies in Appendix G (p. 183), the research questions are addressed in turn using the students’ perspectives.

Student case study

Identity and Aims  Alpha was a student from the Mathematics Department aiming for an upper second class degree classification (a 2:1). She wanted to embark on a career in mathematics research after completing her course and, therefore, considered the mathematical content on her course to be very important. She considered understanding concepts to be important; however, she also appeared to believe that achieving a high score in CAA tests indicated an adequate understanding.

During the first year, Alpha found her studies to be “a little bit harder” than expected, but suggested that it had a positive effect on her determination. It is clear from several parts of the interview that Alpha viewed marks from assignments as indicators of success and performance; and her pursuit of marks was a means of demonstrating the progress that was expected of her.

However, Alpha believed that CAA testing was a contradiction, somewhat, to other assessment in this regard. She noted that they did not always perform well as waypoints
for measuring success: she said that CAA tests had not “been the best way to tell you how you’re doing, or even to properly test you”.

Indeed, Alpha felt that CAA did not sufficiently test students’ understanding of underlying concepts, saying, “I think if the questions for the online test were more about you having to understand certain things and certain concepts, as opposed to just putting numbers into a formula that you find, ...I think that would be more helpful.”

From her interview, it is clear that Alpha’s aims for the CAA tests was to achieve full marks. She said, “I aim for getting 100% on things like that — just because, why shouldn’t I?” Within her mathematics community, Alpha believed it was a commonly-held view that this was achievable, suggesting, “there’s a general consensus that with [CAA], it’s meant to be easier to get higher marks”, noting the “so much feedback”, ample time and opportunity to practise as reasons for this belief.

**External influences**  Alpha talked fondly about the collaboration and support that her colleagues offered to each other. Talking to colleagues in the same course, Alpha discussed her difficulty with grasping proofs and found this to be a widespread problem and there were occasions when students found aspects of the CAA tests difficult.

The library appeared to be a common meeting area for the Mathematics Department students, particular during CAA testing. Alpha had been asked for help while doing practice tests in the library. Alpha had also collaborated with a friend, another student on the same course, and believed that this was part of a supportive and helpful student community. She said, “I know that we try to help each other. At least, with my friends, we try to help each other with explaining things.”

However, while saying that collaboration during the summative testing phase was “cheating”, Alpha had been approached by students to provide assistance. She said, “What was the point, because they’ve got it now in two minutes? They’ve been watching TV while I’ve been sorting this out.”

While discussing her place in the department in relation to her studies, Alpha presented herself as a student that sought independence and wished to avoid resorting to support where possible. She was aware of the specialist mathematics support unit at the institution and that some of her colleagues had attended but she had not. Furthermore, she said that she was initially reluctant to work with other students; only recently had she started working with a friend on the CAA tests. On support from lecturers, she had noted that “most of them are more than happy to spend an hour helping you” and found this surprising, adding, “I found that some of the lecturers have actually been more helpful than you’re led to believe about to do everything on your own.”
From her actions and what they said in the interview, Alpha conveyed a willingness to be somewhat independent of her peers and support services but open to engaging with her lecturers.

**Learning Cycles**  As shown in these learning cycles diagrams (Figure ??), Alpha concentrated on the questions she answered incorrectly in the first instance, working with a colleague more recently and compiling notes that would also assist her in the summative test. Her aim for the practice test phase was to ensure that she was able to answer every question correctly before proceeding to the summative test. In this phase, Alpha used her lecture notes and printed versions of the CAA test as well as a colleague to discuss the feedback with someone. This enabled a continuous cycle of repressing the practice tests and addressing only those questions she had answered incorrectly in the previous test. During the practice test, therefore, Alpha felt she was able to continue on the learning cycle.

However, it is important to note that Alpha placed great importance on conceptual understanding and in ensuring that she was able to carry out procedures without error. Alpha felt that CAA was weak in terms of testing the grasp of concepts — and, therefore, was not effective with regard to this aim — but was effective for refining her procedural capabilities by virtue of the comprehensive feedback offered by the CAA system. Indeed, Alpha had used a technique to obtain feedback from the system that was not wholly intended in its design.

“If there’s a question we don’t get...we go through and answer with ridiculous answers...They all come back ‘wrong’ and you get the feedback for them all.”

Alpha suggested that using this technique was a ‘last resort’ if she could not independently discover how to approach the question. Indeed, Alpha was somewhat dissatisfied that other students used this technique. She said, “I do know of colleagues who just have [this] method that they’ve written down; they just put the numbers in and then that generates an answer. They haven’t really learned anything.”

The summative test, however, appears to highlight some subtle contradictions. Alpha seemed to be satisfied when she achieved full marks, but dropping marks was a cause for frustration. Alpha believed that the marks she received did not always reflect how much she understood. This occurred when she answered a question incorrectly: some questions were worth several marks and required more time and thought, but zero marks would be awarded if the final answer was incorrect.

This would suggest that Alpha placed significant important on the marks she received from the summative test, and for this phase obtaining full marks was her overriding aim.
(A) Alpha aimed to use the first attempt to establish which questions they cannot answer and to print the feedback for these questions.

(B) Alpha used subsequent attempts at the practice test to concentrate on questions she had answered incorrectly. She continued to use the practice tests until she could answer all the questions correctly. Alpha wrote spontaneous notes using the feedback given by the CAA system and these notes would inform Alpha’s future attempts at the questions.

(c) Alpha aimed to achieve 100% in the summative test. She would use the spontaneous notes made during the practice phase during the summative test period. Regardless whether Alpha achieved her goal to achieve 100%, there appeared to be no further path on the learning cycle.

**Figure 5.2:** Learning cycles for Alpha
Alpha seemed to believe that the summative test was not always effective for her aim: she had diligently worked on practice tests to perfect her methods for answering the question and her subsequent errors meant she was not awarded the marks she felt they deserved.

**Effectiveness of CAA**  As the diagram suggests, the learning cycle stops promptly after the summative test. Alpha felt that the summative test served as a conclusion for CAA and that topic: in particular, she was not motivated to examine the errors she made in the summative test. However, it did spur her on in subsequent assessments; she said, “If I get a lower mark...in the tests, it just makes me go, ‘Right, in the written courseworks I’m going to do so much better’.”

It is also important to consider the contradiction between Alpha’s desire to be procedurally adequate and to have a sound conceptual understanding of the topics being tested, and her dependence on the CAA system. As a consequence of becoming acquainted with the system and the techniques passed between colleagues, Alpha developed greater dependence on the feedback that the system provides as a means to answer some questions. Consequently, her development along the learning cycle lacked a progression towards autonomy. This is, perhaps, a reason why Alpha did not progress further in a learning cycle after the summative test; she did not possess the autonomy to diagnose her errors from the feedback or construct new, more challenging learning goals.

The CAA system permitted Alpha to progress along the learning cycle during the practice phase, but her dependence on the feedback and her perception of the importance of the marks proved to be limiting factors for Alpha’s potential to continue on the learning cycle beyond the summative test.

**The student participants**

The participants in the student study were self-selecting individuals that had taken both questionnaires. It was intended to select a broadly stratified sample of students in order to gain an impression of any similarities and differences between and within cohorts. Although the self-selecting sample were not a perfect stratified sample, they offered a reasonable distribution of ability and confidence, as well as their department affiliation. The sample contained representations of four departments and all three categories for degree classification aims.

The sample of nine students are not to be viewed as a representation of the entire student cohort. Rather, they ought to be considered as exemplars. They reported no
additional support needs or difficulties that would prove barriers to being able to use CAA effectively. It is important to note that each student had different histories and identities that would in part determine their activity: indeed, the model for effective assessment attests to the individuality of the learning experience and some of the experiences and influences that these students reported were individual. The issues and influences that affected more than one student in this study offered insights into effects that might affect a greater number of students in the cohort.

During the interviews, the students were asked broadly similar questions (Appendix D, p. 170) relating to their aims, their course of study and their actions and views with regard to CAA. The questions varied slightly between interviews to take into consideration the responses the students gave in the questionnaires.

For the following analysis, the responses the students gave in the interviews have been used to establish each of their paths in the learning cycle while undergoing CAA testing. The students’ learning cycles are compared against those of their peers to establish similarities and differences, with particular regard to their aims and motives, and the departments in which they study.

This was achieved by coding the data with respect to the nodes in Engeström’s (1987) triangular model of activity (48), to the nodes in the model for effective assessment, and to the research questions. In order to preserve the anonymity of the participants, each of the students has been assigned a name — which are letters of the Greek alphabet.

**Students’ identities** The questionnaires and interviews provided an overview of the students’ mathematics qualifications, confidence and affinities. Table 5.2 shows that all the interviewed students believed that understanding all the mathematics would be important and all but one of the sample anticipated that mathematics would be important to their future careers. The interviews identified distinguishing characteristics in members of the sample.
Epsilon, Zeta and Eta were from Engineering Department A and they expressed a high level of confidence in mathematics, whereas Theta and Iota were in other engineering departments and expressed a lack in confidence. Of the mathematics students, Alpha, Beta and Gamma experienced some difficulty with their university course and found the syllabus more challenging than expected.

The theoretical framework chapter referred to the notion of leading identity (Black et al., 2009). It conceives identity as being multifaceted and comprising multiple identities that sometimes complement and sometimes compete against each other. This was an addition to Leont’ev’s (1981) concept of leading identity, which is the prevailing activity that a subject acts within, despite also being associated with other activities. The dominant, leading activity is the manifestation of a dominant leading identity: the activity that prevails is the one most aligned to the subject’s leading identity.

Students are not just students — and this is evident from the interviews. Alpha wished to pursue mathematics research for her career and, thus, becoming a mathematics researcher was her leading identity. For Beta, a potential placement in Australia encouraged her to achieve the best she could, though she did not have specific careers in mind. Her leading identity was driven by the experiences she anticipated in her placement and her leading activity was to do as well as she could in her studies to maximise the likelihood of being able to go.

Gamma had multiple identities: he stated a desire to sample a breadth of university experiences. While he identified as a mathematics student, he was also a volunteer and charity fundraiser with the student union. There was insufficient evidence from the interview that any of these identities would become a leading identity at times of conflict.

Delta was a joint honours student, studying mathematics and computer science, and developed different identities as both mathematics student and computer science student. She found that she did not enjoy her computer science modules and transferred to single honours mathematics, which marked a significant deviation in the actions required to fulfil the object of the activity to gain a degree.

Epsilon, like Gamma, had developed multiple identities at university; he was keen to discover the social aspect in particular, joining some clubs and societies. He also identified as an engineer, and this manifested in his view of mathematics: that it is important to get the correct solution since the outcomes could prove dire “in real life” if incorrect. He also had the motive to be a pilot and this dictated his choice of course and his path through university life.

Theta had previously worked in a field related to the branch of engineering she was studying and wanted to improve her employability. She related her identity to qualities
that she observed in herself: most notably she regarded herself as being independent and determined. These beliefs about herself had a notable influence on her activity, whereby she would persist with the practice tests alone before contacting the mathematics support centre.

Both Iota and Theta believed that their mathematical competence was below that of their peers, and this deficit was subsumed into their identities. They both believed that they had to work harder in order to keep pace with the rest of their cohorts and characterised themselves as ‘underdogs’, though they did not use this term in the interviews. This was somewhat similar to the findings of Reay and Wiliam (1999), who found that students constructed identities from outcomes of assessment and limit their expectations for what they can achieve.

As well as developing their identities as mathematics or engineering students, the participants were also members of the wider university and most — if not all — would have been living in halls of residence and away from home for the first time.

These identities are borne from different objects: students want to be successful in a number of different ways. As well as achieving a target for their degree classification and potential future career, they must navigate social success by forming new friendships. As Gamma and Epsilon demonstrated, students arrive at university knowing that there are many aspects of student life to explore. These activities and their respective identities sometimes cause contradictions.

Students also bring identities with them to university (for example, Chiang & Schmida, 2002; Holdsworth, 2009) and these identities face challenges and transition as students embark on the activity of studying at university. Indeed, many students find the transition to university to be a process of overcoming contradictions and surprises; these experiences lead to developments in their identity (Hernandez-Martinez et al., 2011a).

Thus students’ identities form amid a complex amalgamation of their histories, interpretations and new experiences. As they engage in new activity systems, they learn what is valued in that culture. This is perhaps why students are keen to do well in assessment, since students are awarded marks in exchange for good performance in assessment. Consequently, individuals spend time performing rewarding actions. Brown, Bull, and Pendlebury (2013, p. 7), for example, argued, “Assessment defines what students regard as important, how they spend their time and how they come to see themselves as students and then as graduates. Students take their cues from what is assessed rather than from what lecturers assert is important.”

The importance of assessment and the potential impact on an individual’s identity as a student is evident in the way the students discussed assessment and how assessment
relates to their motives. Without exception, these students worked hard at the CAA tests and used practice tests extensively in preparation for the summative test. Furthermore, many of the participants related the CAA test scores to goals towards their motive to obtain a desired degree classification.

It is embedded within the identities of some of those students interviewed to apply themselves fully to the CAA tests. This is because they regarded their scores as reflections of their identity; scores warranted marks which had an exchange value in terms of their grade at the end of the module. These grades are perceived to be a determinant of what the university judges to be a competent mathematician or engineer and a good student.

This can be observed particularly clearly in Theta’s case. She regarded herself as somewhat separate from many on the course because her aims were not the same as her peers — she wished to pass while others wanted higher scores — and she largely worked on her own:

“Usually they’ll have gone through it once and got 70%. They know how it’s going to be. They know it’s going to be easy.” (Theta, emphasis added)

Theta was not the only case in which she compared her own identity to others based upon their scores and grades. Iota described his identity in terms of being the best he can:

 “[I’m] a bit of an all-rounder, I think, really. I’m at university to get a degree, to get the best degree [classification] I can possibly get. I’m in it for 100% and everything. Less than that will be unacceptable.” (Iota)

Delta and Epsilon were also passionate in their belief that students “don’t deserve the marks that [they’re] getting” (Epsilon) if they cheat or collaborate inappropriately. This suggests that they felt that students were regarded by the marks they are awarded, whether they are merited or not. Consequently, they felt that some students were regarded more highly — or themselves less highly — when others were getting more marks than they should have “rightfully” been awarded.

It is of little surprise, then, that students related their aims in terms of their identity: they wished to attain the marks that they felt befitted their value as mathematicians, engineers and students.
Students’ aims  Although most of the students in the sample expressed some of the difficulties they had during the first year, they had all set high targets for themselves towards the end of the second semester — indeed, all but one student was aiming to achieve a first class degree by the time the second questionnaire was distributed.

The interviews showed that asking about aims for a degree was important: five of the nine students in the study suggested that their aims for the CAA tests were related to the scores they required for their desired degree classification. The questionnaire data suggested that there was no significant correlation between students’ aims for CAA scores and their degree classifications (Appendix F, p. 172).

Some students — and most of the students in the interview study — related their short term goals to their long term goals. The lack of correlation between students’ CAA score aims and their degree aims does not necessarily mean that students do not generally make this connection: the interview data explained that students believed they could achieve higher scores in the CAA tests than in other assessments and viewed CAA as an opportunity to ensure they had a portion of the marks prior to an assessment that may have been a better test of their expectations for a degree classification.

A study reported by Reay and Wiliam (1999) suggested that students relate their future success to assessment performance from a young age; performing badly in a test could “ruin one’s chances” (Reay & Wiliam, 1999, p. 347). The children in the study, aged ten and eleven years, developed identities of themselves and others according to assessment and this, in turn, changed students’ expectations of their performances in future assessments.

It may be a bold extrapolation to suggest that first year undergraduates develop identities and aims in a similar way to primary school children. However, there are parallels between the narratives in the Reay and Wiliam study and the experiences of the students in this study. In particular, there were students that believed that collaboration in a summative test was cheating and would not work with peers during the practice test phase in anticipation of the summative test. They indicated that there was a transition to “more individualised, competitive ways of working, which were increasingly displacing the mutually supportive, collaborative group work to which the children were accustomed” (Reay & Wiliam, 1999, p. 351). It may be that students’ compulsion to aim for scores in CAA may be related to their earlier experiences in high-stakes testing where marks and grades were associated with success.

The Reay and Wiliam study also showed that, even when the students perceived an assessment to be a test of the teacher’s performance, the students viewed the results of an assessment as being a reflection on themselves and thus wanted to achieve the best
they could, particularly given the belief that performing poorly would have a significant impact on their future success.

Students’ views on what makes an assessment effective

The model for effective assessment asserts that an assessment remains effective while it satisfies the goals of the student. Furthermore, to remain being effective, students must be able to progress onto more challenging learning goals and still find the assessment useful with respect to these goals.

Between the questionnaires and the interviews, every student suggested a score they were aiming for in the CAA summative test — whether it was conducted online or on paper. Where students differed most greatly was in the aims for the practice test phase.

Two thirds of the students aimed to be able to answer all the practice test questions before they embarked upon the summative test. Of the other three, one aimed to achieve 90% and the other two — Theta and Iota — were aiming to become more confident.

Aims are important, since the goals that these students aim to achieve have a bearing on the actions the students take. This is exemplified in Beta’s case. She had a new motive in the second semester to acquire sufficient marks to be able to take a placement year in another country. Because she had to acquire those marks, her aim in the second semester was to achieve more in the summative test. She applied herself more in the practice tests to maximise the likelihood of achieving her aim.

The points raised in this section, in addition to the assertion made by Brown et al. (2013) that students take cues from assessment for forming their identities, suggest that these students are also forming their goals to suit the assessment they are given. The students also related success in terms of the marking criteria — whether this was achieving full marks or passing the summative assessment.
Consequently, this results in students setting aims in relation to the assessment, rather than towards their learning. This was also observed by Entwistle and Marton (1994), who noted that in revising for examinations:

“Some students sought to understand only the course material as presented to them, while others, to differing degrees, explored the logical, theoretical, and experiential aspects of the topics they studied.” (Entwistle & Marton, 1994, p. 166)

This affects the scope of the effectiveness of an assessment: if the student is contented to set aims that relate only to the assessment, then there is little scope for developing more challenging learning goals. Indeed, once a student has completed a summative assessment, then the assessment tool would have little or no future use. For the CAA system, students have the opportunity to track their progress through the marks that they receive. These marks communicate a “person’s current status” (Brown et al., 2013, p. 9); thus an improvement in the marks that students receive in the practice test suggest an improvement in the student’s status. However, full marks may communicate a _completeness_ to the student’s status, and improvement is no longer possible.

Indeed, all nine students ceased using the CAA practice tests once the summative test was completed — though some considered using them for revising for their exams. Only one student, Eta, expressed a desire to further develop their understanding beyond the summative test. He saw potential in the CAA tool to deliver questions that tested the depth of his understanding — rather than just at the procedural level — but also felt that time was a prohibiting factor for him to explore and test his conceptual understanding elsewhere. Thus his aim was not feasibly possible.

This echoes a finding made by Henningsen and Stein (1997), suggesting that if students are expected to complete a task in a minimal amount of time and are coerced into finding a correct solution to a question, they resort to procedural rather than conceptual understanding. They said:

“The factors most frequently judged to influence those tasks in which students’ thinking processes declined into the use of procedures without connection to meaning or understanding were the removal of challenging aspects of the tasks, shifts in focus from understanding to the correctness or completeness of the answer, and inappropriate amounts of time allotted to the tasks.” (Henningsen & Stein, 1997, p. 535)
Therefore, for these students and, potentially, to the student culture more generally, scores and marks pervade students’ aims for assessment. Thus, for an assessment to be effective with respect to the students, it must be able to assist students in getting the marks they desire. It should also aid students in forming new goals beyond the marks once they have achieved this goal. The appropriateness of these goals is discussed later in the conclusions of the student study (p. 116).

It is evident from the findings in Entwistle and Marton (1994) that it is not only in CAA that students form specific goals related to assessment rather than learning. However:

“Certain questions did not seem to require personal understanding, simply the reproduction of the lecturer’s understanding, while some students were quite content to leave their understanding at this level.” (Entwistle & Marton, 1994, p. 163)

This implies that there is a potential problem in that students may seek to mimic the understanding of their assessor. Some students in that study found this level of understanding sufficient. Gustin (1985) also suggested that lecturers might be conditioned in their mathematical thinking through prior experiences.

With respect to the model for effective assessment, the CAA tests were only effective while they were appropriate for the students’ learning goals. The students’ aims were limited by their inclination towards goals related to grade outcome, rather than the subject material or level of understanding.

The CAA tests largely achieved this, since the tests themselves gave students an evaluative score on completion. The assessment remained effective until the summative test, since the students were only able to construct learning goals that relate to the outcomes of the assessment. Once the summative assessment was complete, and there was no perceived value in continuing to use the CAA tests: the students would gain no credit for doing so. After this point, the students were no longer creating learning goals, and the CAA system ceased to be effective.

Thus, from the student perspective, an assessment tool may only be effective until they have completed the summative portion of the assessment. The inability or disinclination to construct learning goals that persist beyond the summative test is a substantial limiting factor to the scope of effectiveness of any assessment tool.
Chapter 5  Analysis

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>Γ</th>
<th>Δ</th>
<th>E</th>
<th>Z</th>
<th>H</th>
<th>Θ</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stages in practice test phase</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Made notes using feedback from the CAA system</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Concentrated on incorrect answers</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Engaged with peers</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Practice test actions

Conditions under which the CAA system is effective for the students

Observing the students’ actions more broadly in this practice test phase, there are notable patterns. Every student used the test more than once in order to address errors. No student revised their aim to a simpler goal when taking the practice tests. Furthermore, every student continued in the learning cycle between the practice and summative phases: that is, no student encountered problems in continuing their learning while in the practice test phase.

Nonetheless, actions in the practice phase differed between students and some patterns emerged. Alpha, Gamma and Zeta used their first practice test attempt to obtain feedback from their practice test questions without making a serious attempt at answering the questions. They did this so they could model their future attempts at the questions in the same way as the CAA feedback. Epsilon and Eta had two stages to their practice regime in which one stage was performed with peers and one independently. In these cases, the students were able to engage with others and reap the rewards of doing so, while also ensuring they were able to tackle the problems independently, too.

Between the mathematics students, the actions performed in the practice test phase are broadly similar: every participant used the feedback to develop their notes; they all saved time by concentrating only on those questions they did not answer correctly in their previous attempt; and they all worked with others.

The extent to which these facts apply differ between the individuals, however. Alpha and Gamma were explicit in their intention to print the feedback in order to model their methods for answering similar questions while other students used the feedback to complement their lecture notes. Beta was reluctant to engage with peers, whereas the other students in mathematics willingly engaged with their peers.

With the mathematics students’ actions being similar and noting the observable similarities between Epsilon’s and Eta’s actions, it is possible that practices are shared between students of the same department. Noting that Zeta’s actions are very similar to Alpha’s and Gamma’s, it is possible that there is some kind of information exchange that occurs...
between students of different departments, or that students receive common advice when approaching the CAA tests.

This is because students belong to many communities when they are at university. Not only are they members of a department, they are also members of societies and residents within halls of residences. They also belong to the wider body of students that share resources, such as the library, mathematics support centre and computer laboratories. Since students are members of various different communities, there is ample opportunity for the exchange of experiences and tools.

The students did not give any clear indication that they were adopting practices from students on other courses. Epsilon noted that he shared halls with other students on the same course; it is possible that fellow halls students from different courses may have exchanged methods this way or that these practices have developed historically and have been ‘passed down’ to students in subsequent cohorts. Such arrangements are what Budge (2006) described as informal mentoring, which can occur spontaneously and lead to improved results. However, it is also important to note that Budge (2006, p. 78) also said, “there is no research on how mentoring actually improves academic achievement”. Such communications and sharing of practice may have happened, but there was no evidence in this study.

At the culmination of the practice test phase, it is noticeable that every student achieved their aims. This suggests not only that the students appreciated that the practice tests were suitable for preparing for the summative test, but also that every student made achievable learning goals during the practice phase and were able to measure their improvement.

However, there is a debate whether the type of learning goal that students are setting for themselves are appropriate, despite them being achievable. In the model for effective assessment, the responsibility for selecting goals lies with the student; however there are roles for peers and lecturers in influencing the student to select appropriate learning goals.

Observing the activity system for a student, one can understand why a student would set their goals in this way. Figure 5.3 shows Epsilon’s activity system. He wished to embark on a future career as a pilot. The following exchange between the interviewer and Epsilon shows that marks and grades were the primary motivation — the pursuit of a good degree classification was the leading activity — and ensuring that he understood the mathematics was important, but less so.

Interviewer: What motivates you when you’re doing the practice tests?
Epsilon: Well, it’s my overall grade at the end of the year, because I obviously want to do well in university and get a good degree out of it. I want to try to get the best I can out of every single module, so if I do not so well in other modules, subjects like maths will bring my average up. And I want to try to get the best I can in each individual module, as well.

Int.: So, is it just the sort of marks and grades that motivate you?

Eps.: Yeah, I would say so. I quite enjoy maths as a subject as well and understanding it, because I think it’s quite fundamental to how everything works. So as well as getting the grades, I think it’s really useful for understanding the subjects, because then you can apply it to real-life situations and other engineering examples, as well. So that motivates me as well.

This is significant when considering the effectiveness of assessment for the student. If the student’s motive is to obtain a good degree classification, then it follows that the student will set goals relating to obtaining marks. The students will be culturally aware that marks are the reward on offer that have exchange value: the marks will contribute towards the degree classification. The pursuit of deeper understanding, or doing further work to expand knowledge beyond the syllabus, has less tangible rewards and is considered less important as a result. Even when students do have goals to deepen their understanding in mind, they conflict with other goals or cannot be achieved under the conditions of broader university study. While these goals remain with the student, they are not acted upon.

Therefore, while an assessment may be effective for the student that is pursuing the marks they desire, lecturers may be uneasy about students approaching assessment in
this manner. If that is the case, then the assessment would not be effective for lecturers
that are keen for students to extend their knowledge through assessment.

It is also of interest to note the use of tools in the practice test phase: every student used
at least one additional resource to help them with the practice tests. Most used lecture
notes or course workbooks. Seven of the students also developed tools in the process
of the assessment: these participants noted the feedback they were given so that they
would be able to refer to these notes in subsequent practice tests and, in some cases, in
the summative test.

Such development of tools is a behaviour that has already been described in cultural-
historical activity theory and Sannino noted that this is a human trait: “Tool making
emerges as the individual has to face the constraints and use the available resources of
the natural environment to survive” (Sannino, 2011, p. 575). Therefore, not only do
students do this as a consequence of being human, but they develop these tools in order
to fulfil the actions they need to accomplish in order to achieve their longer-term motives.

Engeström called the section of his triangular model dealing with tool-making, “produc-
tion”. He said, “there is no activity without the component of production” (Engeström
1987, p.86), but he also referred to production as “consumption of the individual’s abili-
ties and of the means of production” (Engeström 1987, p.79).

Therefore, since individual students are producing tools in the course of their endeavours
to achieve a learning goal, they are engaged in production and fulfilling a necessary
element of human activity. The result is the objective of assessment: the students use
their existing abilities and produce an output that becomes an artefact of their learning
— such as qualifications, certificates or reputation — and the result is evidence that they
have achieved their learning goals.

The other three sections of Engeström’s model include the “community”. The students
in this setting are within a community in which they are expected to work and adhere
to the rules. In return, they are given marks, feedback and support. Within the larger
communities that include departments, societies and halls of residences, some of the
students formed smaller cliques in which to “divide labour” in order to complete the
work. It is within these cliques that Epsilon and Eta, in particular, cooperated with
peers.

Not every student believed it was correct to do this. Some of these students believed that
when they were judged as individuals — as is the case in the summative tests — they
should work as individuals. Gamma’s interpretation of the expectations placed upon
them led them to engage more fully with collaboration.
Chapter 5  Analysis

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>Γ</th>
<th>Δ</th>
<th>E</th>
<th>Z</th>
<th>H</th>
<th>Θ</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engaged in collaboration during the practice tests</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred to work on problems in a group</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred to work on practice tests in a group</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believed collaboration is encouraged in the practice tests</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5: Use of collaboration and peer support

It is clear from the interview responses that some students had inconsistent views of collaboration during the practice test phases: some students indicated that they prefer to work on practice tests in a group yet gave no indication that they did so in the questionnaire, and vice versa. This is further confounded when the students responded in the questionnaire in which they were asked whether they believed collaboration is encouraged or discouraged in the practice tests (or that they were unsure). Not only were answers inconsistent within departments — Alpha believed it was encouraged and Delta thought they were discouraged, for example — they also suggest inconsistencies with the students’ actions.

For example, Delta and Eta believed collaboration was discouraged but did so anyway. This may be because they felt collaboration was widespread among the cohort anyway: certainly Eta stated that this was the case. This is an interesting example of a contradiction within a community. Here, some students have interpreted an implied rule on the group activity that students should not collaborate during the practice test phase. Some students did not share this interpretation — in fact, some interpreted the opposite — and engaged in collaboration. For those students that believed collaboration was discouraged, they observed that others were acting differently an made the choice between working independently according to the perceived rules, but perhaps to their own detriment, or follow the lead taken by others and adhere to the interpretation of the rules observed by the majority. In this case, Delta and Eta opted for the latter option.

This suggests that the community created by the cohort of students — and cliques within those — can be very influential in determining students’ actions. As noted previously, there are patterns in the ways in which students take the practice tests and some of the students were prepared to collaborate even when they thought it was discouraged by the lecturer.

It reinforces the position that the model for effective assessment takes on the influence of peers and teachers on the potential for an assessment to be effective. The model says that peers and teachers can influence the assessment at three key stages: setting aims, during assessment and providing feedback. The findings suggest that peers can have a
substantial impact on the effectiveness of an assessment during the assessment phase: not only do they provide opportunities for peer assessment, but peers may also facilitate the student’s learning while undertaking an assessment.

This kind of spontaneous informal mentoring (Budge, 2006) enables students to obtain immediate feedback, which is more personal than that offered by the CAA system. Not only is it an advantage to the students that collaborated in this way, it also gives greater scope for the CAA system to be effective. While Theta worked alone and, occasionally, sought help from the mathematics support centre, others like Epsilon and Eta regularly obtained feedback from their peers.

The interviews could not determine with certainty that these additional opportunities for feedback had a longer-lasting advantage for students’ learning. However, Topping (1998) noted:

“Organised, delivered, and monitored with care, [peer assessment] can yield gains in the cognitive, social, affective, transferable skill, and systemic domains that are at least as good as those from staff assessment.” (Topping, 1998, p. 269)

Though Topping warned that more evaluation of peer assessment was needed, if students can readily draw upon the influence of peers while they undertake an assessment, it is possible that assessments could remain effective for longer.

It is not necessarily trivial to define the activity system: for example, the differences in the way CAA is conducted between departments suggest that these students are in different communities; however, they are members of the same institution and the tools and outcomes are similar. Núñez (2009, p. 10) argued that activity systems are nested and that “few studies take advantage of [activity systems] being nested in other activity systems.”

In such a construction, the influences from larger entities (such as a university or a group that regulates university learning) on smaller entities (departments, lecture theatres, individual study) can be observed and considered. Thus, for each student, they are members of several communities.

The interview study showed that students are members of several different communities, which, as Núñez (2009) described, are nested. Furthermore, they do not form a simple hierarchy of size and do not nest in subsets. The interview data also demonstrate that students moved between these communities and formed cliques within these communities:
for example, some students would work in the library and engaged variably with other students in their department.

This construct of communities helps to understand the sometimes-fluid approaches to CAA practice tests. Some students reported changes in their practice over the course of the first year, in which they encountered different lecturer practices, students from different departments and changing learning goals. It also serves to suggest a possible cause of the similarities of practices between different departments, where one might expect students from the same department to engage in somewhat similar activity.

With students moving between communities — for example, working with colleagues in the library to working somewhat individually as a mathematics student — there may arise certain contradictions. Certainly, in the cases of Delta and Eta, the students were under the impression that the rules of the activity discouraged collaboration yet they formed cliques in which collaboration was encouraged. Such contradictions were called “quaternary contradictions” by Engeström (2014). When presented with such contradictions, the students made a choice and, in the cases of Delta and Eta, they chose to align with the cliques they formed and collaborated.

As noted in the theoretical framework chapter, these students may have developed a leading identity (Black et al., 2009) that directed them towards goals and motives that they believe represents success. For these students, getting a desired score was a marker of their interpretation of success. Consequently, if they believed that the cliques they formed were more compatible with their striving for such success, they would be more likely to align with this activity. Both Delta and Eta were keen to achieve a first class degree and aimed to achieve as high as possible in assessments; in CAA they both aimed to achieve 100%.

However, while there are leading identities that appeared to have drawn students towards the pursuit of marks, some students noted other identities that generated contradictions. Both Epsilon and Eta had a desire to develop a deeper understanding; Epsilon expressed a desire to develop further knowledge and understanding in anticipation of a possible future career in aviation. In both their cases, their leading identity, which directed them towards marks and grades to achieve success, dominated their motive and, thus, somewhat appeased this “primary contradiction” (Engeström, 2014).

There are at least three ways in which this kind of resolution could have been made. Either that desire to attain a deeper level of understanding had abated; or it had been sated, by doing extra work to achieve this goal; or the students believed that it had been sated. The interviews did not give conclusive evidence to suggest which of these three (or another) might have provided a clear resolution. Nonetheless, it seemed that the
students were satisfied that marks were the most significant indicator of success in their studies, and achieving the highest marks possible became their motive. In the cases of Epsilon and Eta, it came at the expense of their aims to develop their understanding.

Solomon (2007) observed that students were willing to sacrifice deeper understanding of concepts to perfect the methods they are expected to learn; she encountered several students in a study that were sufficiently happy to learn the procedures and to dismiss any information that they would not be expected to remember for a summative test.

van de Watering, Gijbels, Dochy, and van der Rijt (2008) also summarised a number of studies that showed that students tend to prefer assessment types that they deemed easier to score marks. Interestingly, they found that students do not necessarily prefer assessment types that they scored well in; nonetheless, they concluded that “it is possible that some students prefer written assessment formats because they are used to it, but not because they are good at them” (van de Watering et al., 2008, p. 655).

This may have had an impact on the scope of the effectiveness of CAA testing. If the students perceive marks and grades to be the pinnacle learning, this would affect their actions. In particular, once a summative test has been completed and the marks awarded, it is possible that students envisage no possible future gain from CAA since it would not award any further marks.

Judging from the students’ learning cycles, only one student expressed a desire to develop their understanding further once they had completed the summative test. Eta was already engaged in mathematics and had studied the Further Mathematics A-Level prior to university. He said that he had found some of the CAA regime challenging; he indicated that there were some aspects of the CAA tests that he would return to, given time, beyond the summative assessment.

He was keen to get a first class honours degree and was striving to apply himself more to meet his aim. However, he found that there was a lot to be done and that the time pressures were substantial. As a consequence, he had sacrificed some of the learning he had intended to achieve, including continuing to use the CAA system after the summative test to redress his errors.

Of the others, two intended to use CAA as a revision tool for their exam and six (two thirds) had no intention of using CAA after the summative test. For those students that had no intention of using CAA after the summative test, they perceived no use for it. One hypothesis is that since the summative test signifies the end of a section of material, and no more marks would be awarded by the CAA system for that material, there is no perceived net benefit to working further on CAA. This may also be true for Alpha,
Delta and Theta when they had failed to achieve their aims for the summative test on some occasions, which signified the end of their learning cycles.

Making the CAA system more effective for students

This work offers the opportunity to identify the limiting factors that prevent the system from being effective for longer. Students’ aims appear to be particularly limiting in cases where students cannot or will not select new aims once the summative assessment is complete.

Overcoming this limitation is not straightforward. It is not common practice for a summative assessment to be to offer additional marks once the assessment is complete. The only reasonable or practical way of overcoming this limitation is to encourage students to develop more challenging learning goals.

One way in which this might be achieved is to develop the CAA system so that it trains students to develop learning goals more autonomously. Boud and Falchikov (2007) described the need for students to become “lifelong learners” and to develop the necessary skills to make judgements and decisions without the support of a lecturer or teacher. This kind of adaptation towards a “sustainable assessment” would “meet the needs of the present without compromising the ability of students to meet their own future learning needs” (Boud, 2000, p. 151).

Cultural-historical activity theory highlights the human pursuit towards motives that signify success and survival. Marks have a large cultural significance and represent success in the forms of aptitude, level of understanding, dedication and future employability (Boud & Falchikov, 2007, p. 401). Therefore, encouraging a change in students’ aims towards deeper understanding would require students to believe that developing this understanding would contribute towards their motives. This requires a cultural change in the academic community, where these aims are rewarded, rather than scoring well in a procedure-dependent test.

It is also interesting to note that, from the individual students’ point of view, the department to which they belong did not seem to have a significant effect on whether they achieved their learning goals. However, Theta’s case (Appendix G.8, p. 213) showed that the change in CAA practice between lecturers introduced failures for Theta to achieve learning goals. These failures were because the summative tests introduced unfamiliar contexts to which Theta struggled to adapt.

This is significant, since some students in the study used the feedback to model their solutions. This may have made those students dependent on the context of the question.
No perceived gains following the summative test
No perceived gains following the exam
CAA would not help to develop a deeper understanding
Failing a test (demotivation)

Table 5.6: Causes of the end of the learning cycle

in the same way that Theta had; yet this dependency was only exposed by a change in CAA practice. The reliance on context demonstrated that the CAA system was not sustainable in the manner that Boud (2000) described: Theta was not adequately prepared to use her knowledge beyond the CAA tests.

The implication is that there is a significant challenge to whether the CAA system had been effective at all — particularly in Theta’s case — regardless whether it had satisfied the aims of students and lecturers. It required a change in assessment practice to expose Theta to a more challenging learning goal: that is, to be able to apply the methods learned to unfamiliar contexts.

It reaffirms the onus upon the assessment to encourage students to pursue more challenging learning goals in order for it to remain effective. While the CAA system was effective at testing a student’s ability to perform a method, it lacked the capability to test students’ conceptual understanding. Had the CAA system possessed the aptitude to challenge Theta’s learning in the way that the paper test did, Theta may have realised sooner that her knowledge and understanding could be developed further. In that regard, Theta was able to achieve her aims without developing the understanding that would be required in other contexts. The pursuit of a new aim might have extended the scope of the CAA system’s effectiveness.

5.1.3 Summary

All of the students were able to achieve their learning goals at least most of the time. There were occasions when students failed to achieve these goals. Some students explained that this was because of the marking scheme: they felt they were harshly scored zero when they should have been awarded for some of their knowledge, rather than punished for making mistakes. In one case, the dependence on the context of the questions presented in the CAA tests led to failures when the summative test presented a different context.
Most students were willing to collaborate on the practice tests while others were reluctant to do so. Gamma suggested that he collaborated during the summative test. The students seemed unclear whether collaboration was encouraged or not, but appeared more inclined to collaborate in situations where they believed collaboration was commonplace. Although collaboration, in itself, did not appear to have a significant effect on the scope for students to achieve their goals, some students appeared to gain more from collaboration than others: these students demonstrated they could achieve autonomy and work independently in a summative test after working with others and alone in the practice test phase.

The students tended to set aims for themselves that related to the marks they wished to receive. On the one hand, this motivated them to use the practice tests to ensure they perfected the techniques they needed to know in order to ensure they achieved their aims. However, this meant that the students had no motivation for further study once the summative test had finished (Table 5.6).

Eta expressed an interest in continuing his learning beyond the summative test but felt that the CAA system was not capable of testing him with respect to his conceptual knowledge. This opinion was shared by another student who felt that the tests were purely procedural. This lack of conceptual testing and the pressure on students’ time was one limiting factor on the effectiveness of the CAA system.

However, it is clear that the students were motivated most by the marks they could receive from the CAA tests. For many of the students, the summative test marked the end of the need to use the CAA tests. Some were prepared to revisit the material and use the CAA practice tests in order to revise for their exams, which extended the effectiveness of the system. No student suggested that they revisited the questions they answered incorrectly in the summative test; nonetheless, the lack of comprehensive feedback, which was present in the practice tests, would have prohibited students from being able to compare their solutions in the way they had been accustomed.

With regard to the model, it would seem that the CAA system has been effective for students while they are striving to achieve a certain mark; and through this study it would appear that students develop and share practices that largely help them to achieve those scores. However, while the marks suggest that the students achieved what they wished to achieve through these tests, there was a substantial part of their learning that remained untested. Furthermore, the students were not always able to identify that this was the case and, therefore, were not exposed to the potential goal to address these shortcomings. This was the cause of all-but-one of the cases for the CAA system to cease being effective.
For students that did not achieve full marks, there was potentially more learning that could be achieved, and thus they might have been able to continue on their learning cycles. However, any effort expended to do so is not given the same cultural recognition — in particular, no further marks are given. The lack of feedback following the summative test might also reinforce the idea that there is little to be gained by persevering on that topic. Some students, like Gamma and Eta, indicated a willingness to reflect on their summative tests, but the lack of feedback at this stage of the assessment was a barrier.

There is a pedagogic reason for not disclosing feedback at this stage: disclosing full-worked solutions for summative test questions risks the validity of future tests should the solutions be made available to subsequent cohorts of students. Some students had the option to return to the practice tests to practise questions similar to those they had answered incorrectly; however, they did not.

Nonetheless, this demonstrates that lecturers’ practices have a role in determining the extent to which an assessment is effective. The change of practice between semesters that Theta experienced meant that her learning stopped once she failed a test; however, it did expose Theta to new learning goals that she pursued away from the CAA system. The choice to suppress feedback after the summative test may have communicated an end to the learning scenario. The choice to allow students to complete the test without invigilation may have communicated to students that it is permissible to collaborate with students during the summative test.

Over and above the desire to explore lecturers’ learning cycles while using the CAA system, it follows from the student study that it is important to establish why lecturers made their choices for CAA practice, since they had a profound impact in the way the students acted with the CAA system.

5.2 The lecturer study

The student study already highlighted the difference between lecturers in terms of CAA practice and one of the emerging issues from this study is that lecturer practice has an influence on the effectiveness of an assessment. However, the student study captured neither the reasons for these divergences in practice nor the reasons why lecturers use CAA at all.

This section outlines the findings from a questionnaire that was used to gauge the practices of the lecturers teaching first year students in mathematics and engineering departments at the institution. They exemplify the diversity of practice that had emerged since
the CAA system was introduced. The questionnaires also give an impression of what the lecturers felt towards the CAA system and how it was effective to their needs.

It also contains a case study of one lecturer to demonstrate the analysis performed on the interview data with six of lecturers that responded in the questionnaire study. These six case studies, given in full in Appendix H, (p. 220), are summarised and compared in this section in order to answer the research questions with respect to the lecturers’ perspectives.

The department hosted and gave lecturers access to the CAA system; however, it remained at the professional discretion of the lecturers to decide whether and how the CAA system would be used with their students. Beyond offering this access, the lecturers reported in interviews that there was no further direction or instruction from their departments; this may be due to the fact that the practice of using CAA tests in the mathematics and engineering departments had continued for many years.

Thirteen lecturers from the department were responsible for the teaching of mathematics modules to the first year students in this study. Through email correspondence with these lecturers, four of the thirteen reported that they were not using CAA with their students at the time. The nine lecturers that used the CAA system were approached for this study.

5.2.1 Lecturer questionnaires

A questionnaire (Appendix C, p. 167) was circulated to the nine lecturers that were using the CAA system with their first year students with the purpose to establish how the lecturers use the system, why they use the system and to explore some of the issues that lecturers may have experienced when using the system.

The responses from the questionnaires demonstrated that there was a diverse range of practices adopted by the lecturers (Table 5.7), with four of the nine adopting somewhat unique practices. All the lecturers granted their students access to the practice tests, but it was in the summative test practice that lecturers differed most. One lecturer did not offer a summative test, but allowed access to the CAA practice tests so that students could test their knowledge without being judged on it directly.

One lecturer replaced the CAA summative test with an invigilated paper test. The other lecturers differed in terms of whether the CAA summative test was invigilated or not: three invigilated the test and four did not.

The lecturers were asked about their stances on collaboration for CAA testing. For practice test phases, all lecturers were happy for students to collaborate, but only three
lecturers explicitly encouraged students to collaborate. Four lecturers prohibited collaboration in the summative test by invigilating the test session; the four lecturers that did not invigilate the summative test were against the idea of students collaborating; nonetheless, they did not communicate their wishes for students to work alone to their cohorts.

Table 5.8 indicated that they tended to agree that using CAA freed time, was convenient, provided opportunities and motivation for students to practise and offered immediate feedback; however the quality of the feedback offered to students appeared to be less of a motivator to provide these CAA tests to students. Furthermore, the lecturers disagreed on the ease of setting up these tests.

Some row totals are not equal to nine in Table 5.8: some lecturers opted not to respond to all the questions.

When asked about the bank of questions that are contained within the CAA system, there was no consensus as to whether the questions posed sufficient challenge to the

<table>
<thead>
<tr>
<th></th>
<th>CAA summative test</th>
<th>@c@Invigilated paper test</th>
<th>No test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to practice test</td>
<td>Invigilated</td>
<td>Non-invigilated</td>
<td></td>
</tr>
<tr>
<td>granted more than one week</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>granted before a test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to practice test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>granted less than one week</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>granted before a test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to practice test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>granted after the test</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Access to practice test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>granted after the test</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: Lecturer practices for the use of CAA

Chapter 5 Analysis

I use CAA with students because...

<table>
<thead>
<tr>
<th>Reason</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>it is easy to set CAA tests for my students</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>it was used by a previous lecturer</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>I am encouraged to by the department</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>students receive immediate feedback</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>students receive good quality feedback</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CAA frees up time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>CAA is convenient</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>CAA provides students with opportunities to</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>practise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA provides students with motivation to</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>practise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8: Lecturers’ reasons for using CAA

students (two believed they do, five did not, two neither agreed nor disagreed). However, the lecturers felt that developing questions took too much time (four agreed, five neither agreed nor disagreed) and was too difficult (three agreed, six neither agreed nor disagreed). Only three had attempted to develop questions for the CAA system and only two thought it would be worthwhile doing so. The remaining lecturers had not attempted to write new questions and did not believe it would be worthwhile.

This offered an early glimpse into the contradictions that lecturers faced while using this system. The questions that the CAA system displayed to students were not sufficiently challenging, yet there was no consensus between the lecturers on whether it would be worthwhile to develop more questions.

In a later group of questions in the questionnaire, the lecturers offered some insight as to why the questions might not have offered the challenge that they desired. They were asked the extent to which they agreed that the CAA system was able to test students on their recall, procedural and conceptual ability.

They were most convinced by the statement that the CAA system was able to test their students’ abilities to carry out mathematical procedures and methods (eight agreed, one neither agreed nor disagreed). Five lecturers believed that the CAA questions tested
students’ recall; three disagreed. One lecturer felt that the system tested students’ conceptual knowledge, but three disagreed. The lack of agreement may be due to the differences between the mathematics and engineering questions, or due to diverse practices that the lecturers adopted.

An invigilated test may well be a test of recall if no additional materials are permitted — indeed, the lecturers that strongly agreed that the CAA summative tests were a test of recall also invigilated their tests — whereas a non-invigilated test may have invited students to use their own materials and thus negated the need for recall.

Both banks of questions — for the mathematicians and for the engineers — were abundant with questions aimed at testing the students’ ability to carry out procedures, yet only one lecturer believed that these questions posed sufficient challenge to the students. There was some desire to develop more questions for the system to address the lack of challenge. These data suggest that perhaps this is because these lecturers wished to add the ability to test conceptual ability more routinely in the CAA tests, and at that time the CAA questions could not provide this challenge.

If this is true, the contradiction that lecturers faced was between the desire to test students’ conceptual knowledge, with the same efficiency and immediacy provided by the CAA tests, and the difficulty in setting new questions on the system. In order to resolve this contradiction, the lecturers would have to find a means of testing students conceptual knowledge.

The questionnaire data provided two means by which some of the lecturers have attempted to achieve this. One lecturer used invigilated paper tests to introduce a conceptual dimension to the assessment. Three lecturers had attempted to develop new questions for the system, despite the difficulty and time cost of doing so.

The conclusions from the questionnaire allude to the existence of contradictions within an established practice of using CAA with students. The lecturers reported that the system is convenient and efficient, and provided students with the means and motivation to practice. However, the assessment itself was insufficient to test the breadth of ability that the lecturers wished to test of their students and there are indications within the data of how the lecturers have sought to overcome this problem.

5.2.2 Lecturer interviews

This section first provides a full case study of one lecturer. It describes the identity of this lecturer and suggests how this affected her aims and actions with respect to her use of the CAA system and the decisions she made as she settled in her CAA practice. These
case studies describe for what aims the CAA tests were effective and at what point the CAA tests are no longer effective for their aims.

The six case studies were used to discuss the aims of the lecturers, and what they seek from assessment. This addresses the first research question, which asks what makes an assessment effective with respect to the lecturers. The case studies were also used to assimilate the similarities and differences between lecturers’ practices, and they these similarities and differences exist. They are used to critique the CAA system and judge its effectiveness for the lecturers’ needs. This addresses the second research question: to establish the effectiveness of the CAA system.

The contradictions that lecturers faced, most notably between their pedagogical aims and the capabilities of the CAA system, were points at which the CAA system could no longer be effected. Thus, from the six studies, the limitations of the CAA system and how they might be addressed offer suggestions for the third research question: in other words, how the CAA system may be improved to extend the scope of its effectiveness.

For the system to be effective, it must satisfy the lecturers’ aims for assessment. These aims are sometimes multi-faceted and the assessment can be effective for some of these aims, but not others. If the lecturer has achieved a stable practice, then one can say that the assessment tool has been effective to some extent, but there may be other aims that the tool cannot fulfil.

Six lecturers indicated on their questionnaires that they would be prepared to participate in an interview to discuss CAA further. They lasted between 27 minutes and 54 minutes, with a median time of 34 minutes. They are given Greek letters as pseudonyms: namely Kappa, Lambda, Mu, Nu, Xi and Omicron.

**Lecturer case study**

**Subject** Kappa was an experienced lecturer whose research was in the field of mathematics pedagogy. She reflected upon her teaching practice in light of her own research and her reading of others’ contribution to mathematics pedagogy literature.

She was regarded highly for this expertise both within the department and more widely in the mathematics education community. As Roth (2007b) argued, identity is a construction from what an individual has done and what they think. In Kappa’s case, her research, her practice and the beliefs that she had generated through her work comprised her identity and moulded the way she acted when teaching.

Using the notion of figured worlds (Holland et al., 1998a), Kappa had developed identities as a teacher, researcher of mathematics pedagogy and expert in mathematics education.
These figured worlds were by no means disparate: they were related and were mutually influential. For the activity of assessing her students, Kappa spoke from each of these identities, demonstrating the influences that they have on each other: in particular, she noted the introduction of the group projects as a result of a research project she had been working on.

**Object** Kappa’s object for the activity was driven by her research and reading. She emphasised the need to allow students to explore the concepts behind the mathematics they were using, which needed to occur alongside giving students the confidence to perform the procedures they are required to learn. Therefore, she realised the importance of giving students the opportunity to practise and gain feedback on their work.

By offering regular interim assessments, Kappa could monitor students and allow students to monitor themselves in their learning. However, such regular testing would be a burden on her time in preparing and writing meaningful feedback using traditional assessments techniques. While she wanted to offer students this facility, she acknowledged that she was already pressured for time.

**Tools** She used CAA with her students in Engineering Department E. In the past she had been reluctant to use some online assessment resources as she found them to lack rigour and had been rudimentary. She had been generally keen to introduce technology in her teaching where it would help students to grasp topics and develop their conceptual awareness.
She used the CAA system with these students since they offered the opportunity for students to practise: “I think that they fulfil a very useful purpose in enabling students to practise and become familiar and confident with the way of doing things.”

This was not the only aim that Kappa had for the interim assessments. As well as offering the opportunity to practise, Kappa desired to test students’ conceptual understanding: “I put an important emphasis on conceptual work, but that doesn’t say there’s no need to practise.” Kappa felt that the CAA system was not sufficient for testing conceptual understanding in its current state, but fulfilled the purpose of ensuring that students were tested on their procedural work.

For these reasons, Kappa found it insufficient to depend on one tool to carry out the activity and adopted other assessment techniques to reach the object of the activity. While CAA was a useful tool for testing many students quickly, other assessment tools could be used to test conceptual understanding. She said, “If they [the CAA tests] were the only form of assessment, then I would feel that maths learning was impoverished, because there is a lot of encouragement for instrumental learning.”

Kappa was keen not to reduce the assessment regime to one kind of assessment tool; she was content to have the variety of CAA alongside projects to assess different types of knowledge of understanding. For that reason, she was not keen on the idea of introducing questions that assess conceptual understanding to CAA: “It’s not so much that I want to put more challenging questions on the CAA, but by themselves, they [the CAA tests] don’t offer so much of a challenge as I want. But given that the CAA is part of a wider assessment programme, then the CAA alone is not offering the degree of challenge I want overall.”

Community  Kappa was a member of many communities that had an impact on her activity of assessing her students. She was a member of the mathematics lecturers and department; she was a member of local, regional and international communities of mathematics education researchers; and she keenly kept abreast of innovations in mathematics learning, pedagogy and assessment literature.

These communities affected her in many ways: not least in the formation of her leading identity for this activity. She was aware of the various tools at her disposal that would help with her activity: from the CAA made available by the department, to the techniques discussed in the other communities.

Division of labour  One of the many reasons that lecturers are tempted to use the CAA system is the opportunity to save time by having assessments marked automatically
by a computer. Further time is saved in the preparation of such assessments since a staff member is dedicated to setting up tests to the lecturer’s liking. In practice, this meant that lecturers could request which questions appear in the tests, how long the practice tests are made available, and precisely when the summative test is made available.

Kappa made use of this division of labour: “What’s particularly good about them…is that more or less somebody else does all the work. So, I get a form of assessment that is useful within the boundaries that I’ve stated, and it happens. Whereas for the projects [which assess conceptual understanding], I’ve just spent almost the entire weekend marking a set of projects…They are hugely labour-intensive, whereas CAA tests are not.”

The workbooks, which are given to students as a core document for each topic within their engineering mathematics modules, also offered the students the opportunity to explore the knowledge they had acquired away from lectures. They allowed the students to explore different contexts in which to use their mathematical knowledge.

**Rules** Like all the other lecturers, Kappa was bound by the rules of the organisation regarding the use of assessment, including safeguarding against the effects of plagiarism. Although she was compelled to act upon plagiarism incidents, she was intent on avoiding such instances by ensuring that the students could not collaborate during summative assessments. To that end, she invigilated the CAA summative tests.

Invigilating the test also provided benefits in terms of the trust Kappa could place in the scores students gained. When Kappa was asked why she invigilated the summative test, she said, “How do you know who’s done it? How do you know that they’ve done it on their own? How do you know if they’ve copied from somebody else or from the book? I invigilate because I want to see what each student working on their own can achieve.”

She was also a member of a wider mathematics education research community, which had an effect on her identity as a lecturer and educator, the object of her activity and what she considered to be good assessment practice. The principles of this good assessment practice manifested as a set of rules by which she was obligated to follow in her assessment regime, which included:

- the need to assess students on their conceptual understanding as well as their procedural capabilities;
- to monitor her students’ progress, and allow students to do the same;
- to offer regular and timely feedback.
The influence of the rules by which she abided had a clear influence on the object of her activity. These rules were beyond the expectations that were placed on her by the department. She had noted particularly the need to test the students on their conceptual knowledge during the semester; she found that the CAA system was not capable of testing students to the extent she required.

Since the tool satisfied many of her aims and brought her closer to the object of the activity, the contradiction was not sufficient to warrant abandoning the tool. The CAA system was also considered a part of the learning regime among the engineering department, since it operated as an adjunct to the workbooks.

Contradictions There were two main sites contradiction that Kappa experienced in the course of assessing her students: a primary contradiction that formed between the identities she had formed within her communities, and a secondary contradiction between the sole use of the CAA system and the object of her activity (Figure H.1).

Kappa developed a leading identity as an assessor, which incorporated aspects of each of the lecturer, researcher and expert identities that informed her practice. One effect of complementing these identities was that Kappa remained critically reflective with her practice and was perhaps the least influenced of all the interviewed lecturers by learning experiences at school and university. Gustin (1985) suggested that mathematics experts might hold beliefs influenced by these early experiences, but Kappa was a counter-example.

However, there emerged conflicts between identities. Kappa was keen to develop her assessment practice by incorporating technology — using geometry software and having a weekly session for her students in a computer laboratory, for example — but this had an impact on the time she spent on her teaching. Each of the activities associated with these identities demanded much of her time, and she had to deal with the conflicts arising between her identities to manage the time she spent on these activities.

Inevitably, she felt that she could not develop her practice to exactly how she would want to have it had she had the time to do so. She acknowledged that she had to make compromises between her ideals, which were motivated by her identities. Nonetheless, she was satisfied with her practice and it had reached a stable state.

Reaching this stable state relates to the notion of the self as a leading activity, since Kappa exemplified the development of a “real-life activity that most explicitly positions individuals to meaningfully contribute to the ongoing social collaborative practices in the world” (Stetsenko & Arievitch, 2004, p. 493). That is, Kappa used her experience as a lecturer, mathematics pedagogy researcher and mathematics education expert to
Kappa faced a secondary contradiction between the object of the activity and the CAA system tool that she was using. As noted previously, Kappa’s interpretation of the rules influenced the object of the activity. Like some of the other lecturers in the study, this rendered the CAA system insufficient for achieving for satisfying the object of the activity.

In Kappa’s case, the need to assess students on their conceptual understanding was not satisfied by the CAA system. Although she was not keen to add conceptually challenging questions to the CAA, Kappa noted that one of the barriers that she faced when using this system was that she could not design her own questions for the system. This meant that she relied on the existing question bank. She also felt that the CAA tests “encourage a more instrumental way of seeing mathematics” that reduces mathematics to a set of procedures to follow.

She addressed this contradiction by introducing group project work. In doing so, she was able to challenge the students’ conceptual understanding and to gain experience of new contexts. Although this was a recent addition to the course, Kappa was content that the projects had served the purpose for which they were introduced. That is, they resolved the contradiction that arose between the object of her activity and the tool.

The introduction of group project work created a new contradiction: it proved to be a burden on her time, which the CAA system promised to save. However, Kappa was satisfied that the time saved by offering CAA tests mitigated against the additional time expenditure for marking the projects.
Effectiveness  For Kappa, the CAA system was effective for the testing of students’ individual work on procedures and it satisfied her aims to assess her students quickly and obtain reliable indicators of their performances. Her CAA practice had developed over several years, so at the time of the interview, her practice using the CAA system had become stable.

She reported that the last major change to her practice was to reduce the number of CAA tests from four to two over the course of the semester, and to make the tests twice as long to accommodate the testing of the material. She reasoned, “With the four CAA tests that I had, many students were finishing very quickly. So, being given twice the number of questions didn’t seem unreasonable.”

While CAA remained effective for Kappa’s goals to assess students’ procedural ability, she could not see a use of CAA as a means to accomplish her goal to assess students’ conceptual understanding (Figure H.2) and, this, the CAA system was not effective for this aim. Although this did not cause problems for Kappa — she was satisfied in using other assessment types for this aim — it served to illustrate one of the limitations of the CAA system.

Lecturers’ views on what makes an assessment effective

The model for effective assessment positions lecturers’ aims as an intrinsic measure of whether an assessment has been effective or not. Therefore, it is imperative to study their aims. Furthermore, motives drive agents to act within an activity system and, thus, the lecturers will act according to their aims. Thus this study examined how varied the lecturers’ aims were and whether these differing aims were the cause of their diverse CAA practices identified in the questionnaire study with lecturers.

The lecturers’ paths through the learning cycle suggest that there were many differences between individual lecturers’ aims and outcomes when using the CAA system. With such stark differences, the result is a variety of practices. This is because the CAA system is the tool that the lecturers used to perform assessment and the lecturers adapted their use of the tool as they faced contradictions between what they aimed to achieve and what the CAA system could offer.

For example, both Kappa and Lambda felt that the CAA tests alone were not sufficient as a test of students’ conceptual knowledge and encouraged students to adopt an instrumental way of learning. Their responses to these problems were different, however. Kappa reduced the number of tests and introduced projects to help assess conceptual understanding and broaden the students’ experiences of the contexts in which problems can
occur. Lambda conducted a paper test that introduced real-life contexts and additional, optional coursework that students were free to attempt at their discretion.

In the cases of Kappa and Lambda, their aims for CAA were sufficiently similar to arrive at the same contradictions; however, because they had other aims to satisfy, their response to the contradictions was markedly different. In Kappa’s case, there was a particular need to keep the workload from marking low and to encourage students to collaborate: the projects were a necessary addition to the assessment scheme and the CAA tests retained the ability to offer feedback. For Lambda, marking the tests individually allowed him to maintain a “human” element to assessment, which he viewed as important.

Therefore it is important to analyse the lecturers’ aims for assessment more closely and explore the journeys that the lecturers have made in their learning cycles: in particular, how they overcame the contradictions that they faced to reach their current practices.

Table 5.9 gives an overview of the aims that the lecturers expressed with respect to CAA. All the lecturers were keen to provide students with access to a means of practising their newly acquired mathematical knowledge. Indeed, the desire for students to “engage more in their own learning” (Green et al., 2004, p. 2) was a motivation for the design of the CAA system.

The aims also reflect some of the aims imposed by the university — the community in which the lecturers engaged in the activity of teaching and assessing. Providing a fair means of assessment is an imposition from the community upon the lecturers’ aims. While all the lecturers were required to provide a fair assessment, only Xi mentioned this as one of her aims. This may be because she felt that was an important aim for her, personally.

Some lecturers were happy to offer access to the CAA practice tests, with some keen to encourage students to work between lectures. Although none of the lecturers stated explicitly that they felt that students would not practise between lectures in the absence of assessment, the researcher notes that this sentiment was expressed anecdotally and informally by lecturers on more than one occasion.

In encouraging students to practise between lectures, the lecturers wished for students to consolidate their learning, to increase their confidence, and to be prepared for building upon their learning in the next lecture. This sentiment was echoed by Nicol and Macfarlane-Dick:
<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Aims</th>
</tr>
</thead>
</table>
| Kappa    | • Enable students to practise  
           • Test conceptual understanding  
           • Test students quickly and save time |
| Lambda   | • Encourage students to work between lectures  
           • Enable students to practise  
           • Differentiate between students  
           • Prepare students for a variety of contexts |
| Mu       | • Test a large number of students quickly and efficiently  
           • Establish the extent to which the students have grasped the content  
           • Test students on their ability to carry out procedures  
           • Test whether students are capable beyond the lecture content  
           • Enable students to practise |
| Nu       | • Encourage students to work between lectures  
           • Expand and develop students’ views of what “doing mathematics” is  
           • Consolidate learning that happens in lectures |
| Xi       | • Provide a fair means of assessment  
           • Enable students to practise  
           • Encourage students that are struggling with the content |
| Omicron  | • Allow students to monitor their progress  
           • Enable students to practise  
           • Provide a convenient form of assessment for students |

*Table 5.9: Lecturers’ aims for assessment*
“One effective way to develop self-regulation in students is to provide them with opportunities to practise regulating aspects of their own learning and to reflect on that practice.” (Nicol & Macfarlane-Dick, 2007, p.207)

One reason why they may feel the need for students to practise between lectures is through their own experiences of learning mathematics. Gustin (1985) performed a study to look for similarities between leading American mathematics researchers. In interviews conducted with these researchers and their parents, Gustin noted:

“Perhaps the most significant aspect of their formal secondary school experience was that it did not disturb the tendency of the mathematicians to be independent learners.” (Gustin, 1985, p. 302)

In comparing secondary school education to the mathematicians’ experiences in college, he added:

“For many of the mathematicians, college was the first opportunity to be with other students who had similar interests and abilities...Whereas they had previously learned much of the mathematics they knew on their own, in college they enjoyed competing with and learning from peers. They still studied and worked on their own, however.” (Gustin, 1985, p. 320)

The mathematics researchers in Gustin’s study were used to developing and honing their mastery of mathematics on their own. It may be that the lecturers in this study had similar experiences to those in Gustin’s study, and were therefore inclined to learn independently and engage with like-minded peers. If they believed that this was the right way to learn mathematics, this would explain why they were keen for their students to develop similar habits.

Certainly, the description that Omicron gave of how students should collaborate — “I think, for the students to collaborate on things, it’s helpful for everyone. The strong students learn it better if they explain it and the weaker students learn it better if they’ve had it explained more times” — was similar to the exposition of the mathematicians in Gustin’s study. However, the interviews in this study did not examine the lecturers’ prior experiences of learning mathematics, so comparisons between their approaches to learning and how they teach cannot be made here.

Some of the lecturers used CAA in order to save time in assessment. Kappa, Mu and Xi, in particular, noted that giving feedback for an assessment could be time-consuming,
particularly for longer coursework in which students are expected to express their conceptual understanding. Mu noted that ideally feedback would comprise “me sitting in a room with each individual”, but that would not be realistic. The feedback that the CAA system offered was static, but it was timely. This proved to be a sufficient mitigator given that there were alternative sources of feedback, such as through their tutors.

Mu also wished to establish that students had satisfactorily developed the ability to carry out certain procedures. Therefore, Mu’s aims for CAA was to assess students on their procedural knowledge in a timely and efficient manner.

Other lecturers expressed aims for assessment that were not necessarily realised by using the CAA system alone: Lambda and Nu wished to encourage students to be able to apply their knowledge freely to a variety of contexts; Kappa and Mu wished to test students beyond the procedural content of the course; and Xi and Omicron aimed to provide a means of assessment that was fair and accessible to their students.

Conditions under which the CAA system is effective for the lecturers

The interviews gave evidence that the lecturers had multiple identities, which had an effect on the way they used the CAA system. As well as being mathematics lecturers, they were also mathematicians. Furthermore, there were expectations placed upon them as researchers or experts in their fields. These identities sometimes caused primary contradictions (Engeström & Sannino, 2010) within the subject of the activity. Therefore it is important to discuss multiple identities and their potential impact on activity.

Stetsenko and Arievitch paraphrased Leont’ev’s (1978) belief that “the human self has no history, and no logic of functioning and developing, beyond the history and logic of functioning and developing of human practical purposeful activity” (Stetsenko & Arievitch, 2004, p. 484). That is, one can only make deductions about a person’s identity by observing their actions.

Roth (2007a, p. 56, emphasis in original) gave the following example: “We do not know who a person is independent of the actions of that person. Being shown the middle finger by another car driver, we may think, ‘he is a rude person’. . . Attributions about how who someone is are made based on observable behaviour (actions).” Here, Roth noticed that there is historicity to human actions that help others to develop an impression of that person’s identity.

Therefore, there is a social and historical importance attached to actions: once an individual performs an action, others develop an impression of that person that persist after
the action has been completed. There remains a question of the effect that perceptions of identity formed by others has on the actions of that individual.

Nkomo and Cox Jr (1999) wrote about social identity theory: particularly noting that an individual’s notion of their own identity can be influenced by external factors. They suggested that individuals respond to the expectations that are put on them, even if they do not consciously accept this as part of their identity. They gave an example of gender: individuals may act within expectations for their gender if they do not necessarily regard their gender as important to their identity.

This supports the idea that individuals may act in accordance to expectations placed upon them, even if they do not readily share these expectations in the interview: particularly since they may not be consciously observed. Kappa had expectations placed on her as a leading researcher in mathematics education, yet she did not mention this in her interview.

It is not a straightforward task to offer an objective interpretation of the impact of identity on activity, since not all of the many facets of a person’s identity will be discussed in an interview. Furthermore, as Nkomo and Cox Jr (1999) suggest, the individual may not be aware enough to be able to share all that is relevant to the study.

Identity is important, however. Reay and Wiliam (1999) suggested that there was a bilateral link between identity development and activity: in their study, children spoke about themselves as a consequence of their activity (“I’ll be a nothing” refers to a child that used this phrase to describe her prospects after scoring poorly in assessment).

While past activity has an effect on identity development, Reay and Wiliam noted that identity development resulted in changes in activity: social activity in particular. They noticed that the assumption of different identities within the classroom created a hierarchy of ‘cleverness’, resulting in “more individualised, competitive ways of working, which were increasingly displacing the mutually supportive, collaborative group work to which the children were accustomed” (Reay & Wiliam, 1999, p. 351).

Kappa faced multiple identities that challenged the way that she acted. Her activities associated with her identities competed for time; therefore, she was often addressing tertiary contradictions (Engeström, 1987b) between activities. Her identities as researcher, expert and lecturer were partly constructed from her activities, but also as a result of the interpretation of her activity from the communities to which she had become known. In particular, her reputation as a mathematics education expert was initiated by her activity as a researcher, but developed as her work became known in the mathematics
education research community. This duality between activity-developed (internal) identity and socially-developed (external) identity places internal and external expectations on the way the individual acts (Nkomo & Cox Jr, 1999).

She was keen to develop her course materials, to make them more technology-based and interactive. However, she found that this would impact on her ability to carry out her other activities, which were related to other identities. In resolving these contradictions, she ended up developing what she could but accepting that she could not realistically achieve everything that she would desire.

With regard to CAA, Kappa accepted it as a tool that satisfied most of her aims but could not fulfil all aspects of assessment that she required. She felt that it was not as robust or as challenging as was needed, but in the absence of time it was satisfactory.

Other lecturers had similar experiences of tertiary contradictions with other academic activities. Mu was a successful mathematics education researcher and a successful mathematics lecturer; she used her research to inform her teaching, and her teaching to inform her research. She felt that the system, in its current state, helped to free up her time so that she could spend more time on other aspects of her teaching or her research; developing the system would require considerable time that she could not spare.

Nu had experience of developing the system, and found that it took a substantial amount of time to do so. Having made those changes, he felt that the reward was insufficient to warrant further improvements. Therefore, he offered alternative assessments and individual feedback to those students that requested them; he found this to be less time-consuming and more rewarding.

The interviews also provide evidence of the lecturers’ identity as a lecturer being a motivation to develop practice. Lambda and Xi gave notable examples of how their identities as lecturer gave them a sense of responsibility towards their students to do their best to provide equitable and useful assessment. They both drew upon their previous experiences — Xi, most notably, as a secondary school teacher — to empathise with their students and to critically examine their own practice.

While the lecturers had reached a stable practice of using the CAA system with their students, it was clear from the interviews that they continued to face contradictions between their aims for assessment and what could be achieved from using the CAA system. These contradictions became the limiting factor for the CAA system to remain effective. Making the assessment effective for more of the lecturers’ aims would require addressing these contradictions.
Making the CAA more effective for lecturers

One might be tempted to believe that lecturers have greater power to achieve their aims through assessment than a student would: after all, they decide on the type of assessment, the assessment tool, the timescale of delivery, the marking, the questions that appear in the assessment, and so on; whereas the student merely receives an assessment and is expected to complete it on the instruction of his or her lecturer.

However, this section highlights the fact that lecturers face contradictions when using assessment and there are constraints that limit the freedoms and power they have when offering assessment to their students.

Feedback A common contradiction among the lecturers’ responses was that the feedback provided by the CAA system was static and impersonal. Lambda, Mu, Nu and Xi felt that the feedback offered to students ought to be more individual and tailored to the students’ responses. Nu believed that because the feedback was so static and unresponsive, it could not be considered to be true feedback. Lambda believed that this was not realistically achievable without a human providing feedback.

The lecturers responded to this contradiction in different ways.

- In Lambda’s case, the practice tests offered static feedback, but he was able to provide individual feedback in the summative test that he designed to be performed by hand.

- Mu felt that the feedback was adequate for the level of understanding that the CAA system was testing at; that is, for assessing at the procedural level of knowledge, the feedback offered students the direction they needed to grasp the methods required of them.

- Nu asked students that wished to obtain additional feedback to approach him.

- For Xi, this contradiction remained somewhat unresolved: she was aware that the feedback did not always help those students that were struggling with the material, but had not changed her practice.

The feedback, alone, was responsible for divergent practices. More than half of the lecturers were satisfied enough to continue to use the feedback in its current state; Mu and Xi were less satisfied but continued this practice nonetheless; Lambda and Nu both offered opportunities for more detailed feedback.
## Chapter 5 Analysis

<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Contradictions faced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kappa</strong></td>
<td></td>
</tr>
</tbody>
</table>
  - Wanted to encourage more conceptual understanding; felt CAA encouraged instrumental learning  
  - Used CAA to save time with assessment; also using projects to assess conceptual understanding, which are time-consuming |
| **Lambda** |  
  - Wanted to encourage flexibility and dexterity in applying mathematical knowledge; felt CAA encouraged context-dependency  
  - Believed a human aspect was necessary in assessment; CAA removed the human aspect from assessment |
| **Mu** |  
  - Wanted to give students individual feedback; CAA offered static feedback  
  - Was aware of potential cheating; cohort too large to invigilate the tests  
  - Wanted to encourage students to practise; CAA encouraged students to practise too much |
| **Nu** |  
  - Was aware of potential cheating; cohort too large to invigilate the tests  
  - Wanted to “change [the students’] view of what it means to do mathematics”; felt CAA encouraged context-dependency  
  - Wanted to give feedback that responds to the student’s work; CAA feedback was static  
  - Wanted to limit the number of attempts students made; CAA system did not adequately handle this change  
  - Wanted to ask more challenging questions; CAA was limited in types of questions it could ask |
| **Xi** |  
  - Wanted to encourage students, particularly the weaker students; CAA marking was harsh and demotivating  
  - Wanted to direct students towards improvement; felt the CAA feedback was not helpful to the weaker students |
| **Omicron** |  
  - Wanted to develop new questions for the system; it was time-consuming and onerous to write new questions  
  - Was keen for students to practise while using the system; felt the current bank of questions were not the most effective for this task |

**Table 5.10:** Contradictions that lecturers faced when using CAA
Questions in the CAA system There were further contradictions with some of the questions that the CAA system presented to students. Kappa felt that, overall, the questions were not sufficiently challenging; Lambda felt that the questions encouraged dependence on a narrow range of contexts; Nu felt that the questions lacked significant challenge to his students; and Omicron was unsatisfied with the question bank and wanted to write new questions for the system.

As with the feedback, the lecturers responded to their issues with the CAA system questions in divergent ways.

- Kappa dealt with the insufficient test of conceptual understanding by introducing group projects that permitted students to explore a range of contexts and apply their conceptual knowledge.
- Lambda used the summative tests as a means to introduce new contexts to the methods that students were expected to be able to use. This, in turn, helped to prepare students for the exam and facing unfamiliar contexts. He also offered students additional, optional practice questions.
- Nu offered students additional tutorial questions to introduce new concepts. He wrote some new questions for the CAA system question bank.
- Omicron attempted to write new questions for the question bank, but found this was not a satisfactory way to overcome the contradiction. She enquired about introducing a new CAA system that was able to handle new question types.

Kappa, Lambda and Nu responded to the contradictions they faced with regard to the questions by offering additional means for students to explore their conceptual understanding and contextual dependency. This response also triggered a further contradiction: the CAA system was intended to save lecturers’ time, yet they were faced with having to spend more time addressing the shortcomings of the questions.

In Kappa’s case, she was satisfied to use an additional form of assessment. On the one hand, the CAA system was useful in saving time in testing students’ procedural ability; on the other hand, project work allowed students to work in groups and develop their conceptual understanding. She believed that the time spent on marking the projects was offset by the time saved in offering the CAA tests.

In Lambda’s case, offering a paper summative test fulfilled many of his aims for assessment. Not only did it allow him to expand the students’ experience of the real-life contexts in which the mathematical methods could be used, it also allowed him to retain
the human aspect of assessment that he believed was necessary. Therefore, he felt that the time spent in this exercise was necessary.

In Nu’s case, few students had approached him for feedback. Thus offering this opportunity for additional feedback did not have a significant impact on his time.

For Omicron, the contradiction remained somewhat unresolved. Although she had attempted to develop new questions for the system, this proved to be an ineffective response to the problem. She continued to search for alternative systems that would ask the type of question that she wished to pose to students.

The students’ approach to CAA  The interviews highlighted a potential problem in that the lecturers tended to be unaware of how the students conducted the tests. For some lecturers, this was a concern in those tests that were not invigilated. Only Xi appeared to have an awareness of how students approached the tests; otherwise, the lecturers could only speculate.

Because of the lack of awareness in how students approached these tests, many of the lecturers expressed some concern as to how students approached the tests and whether some may be cheating unbeknownst to them.

Lambda felt that not invigilating the summative test reduced his trust in the results. Mu and Nu were unable to invigilate the summative tests since the cohort was so large; therefore, they allowed the students to conduct the test without supervision. They were aware that cheating may have taken place. Xi had both experienced cheating in the summative tests, while Nu and Omicron had suspicions of cheating.

These lecturers responded to this contradiction in the following ways.

- Kappa invigilated the summative test to ensure that she could trust the CAA scores given to students.
- Lambda invigilated the summative test to ensure that he could trust the CAA scores given to students.
- For Mu, the CAA scores did not have a significant contribution towards the students’ module results.
- Nu also felt that the CAA scores did not have a significant effect on the students’ module scores.
- Xi invigilated the test and was more proactive than the other lecturers in monitoring student activity by examining their data logs collected by the CAA system.
The issue of cheating in assessment is important to consider. Knight (2002) argued that the reliability of an assessment is in part determined by the correlation between evidence and criteria: that is, the assessment must be an adequate test of the criteria for success; and likewise the scores given to students in an assessment must indicate how well the students have performed in relation to those criteria. When that is not the case, the “conclusions are much less reliable, because the goodness of fit between evidence and criteria has to be inferred” (Knight, 2002, p. 277).

Cheating, then, erodes the trust that assessors — in this case, the lecturers — have in the relationship between students’ scores and the extent to which the students have achieved the assessment criteria. Omicron commented that it is a requirement to assign students a score based upon their performance; thus if these scores are not reliable measures of student performance, then the assessment fails to achieve this basic aim.

The disparity between scores and achievement, therefore, creates a contradiction against the lecturers’ aims to give an indicative and reliable score of their students’ achievements.

It is telling that the lecturers responded to this contradiction in different ways. In the cases of Mu, Nu and Omicron, the opportunities for students to cheat were more prevalent since the test was not invigilated; however, they believed that the CAA scores bore such little significance to the module marks and the degree classifications students received that no further action was necessary. Kappa, Lambda and Xi had decided to invigilate their tests in order to maintain the integrity of the summative test.

The split between the two groups of lecturers with regard to their response to the cheating problem also aligns with whether the cohort they were teaching mathematics students or engineering students: those teaching mathematics students tended to be less concerned by cheating than those teaching the engineering cohorts. Furthermore, the CAA system between the two groups was subtly different insomuch that the question banks were different between the mathematics and the engineering departments.

The engineering bank of questions was developed as part of a project that also included compiling workbooks; these workbooks were used in many of the engineering mathematics modules. Using these resources helped to create consistency between modules; the CAA system was seen as integral part to the teaching and assessment in the module courses. Therefore, the lecturers’ responses may have been influenced by the perceived need to maintain existing practice.

While one lecturer had evidence of students cheating and another lecturer had strong suspicions that cheating was taking place, a particular problem that led to the lecturers’ questioning of their trust in the CAA system was that they were broadly unaware how the students were undertaking the CAA tests.
Student collaboration  Within the contradictions of the students’ approaches to undertaking the CAA tests was the extent and ways in which the students collaborated in both the practice and summative phases of the CAA testing regime.

The interviews identified that the lecturers tended to have preferences or suggestions as to how the students can engage with collaboration to maximise their learning. However, both the interviews and the questionnaires indicated that lecturers had not shared these preferences and suggestions with their students. Certainly, most lecturers were willing for their students to collaborate in their mathematics modules as a means to discuss the mathematics and to help grasp the methods required; however, this did not extend to policy or advice for students to follow when conducting CAA tests.

Indeed, during the interview, Nu identified that there was a “grey area” in terms of the plagiarism policy, since the policy only appeared to apply to written work, yet he could give feasible examples of student actions that he would consider cheating that related to improper collaborations: “They can have a group of five standing around while they are taking the coursework test and they they could move onto the next person. I have no way of knowing whether that happens. I think that, clearly, would be a violation of the coursework policy.”

This illustrates the problem that lecturers had: they would be encouraging of collaboration between their students, but they would be unaware how those collaborations worked in practice. Some lecturers suggested how students should collaborate but appeared to have little confidence that students actually worked in this manner. As Nu pointed out, there are ways in which collaboration could be construed as plagiarism.

While invigilating the test removed the problem of plagiarism affecting the students’ test scores, the problem that students might not have been learning effectively by engaging in ineffective collaborations remained. Omicron, in particular, felt that there would continue to be a disparity between those that grasped the material quickly and those that required more explanations if the weaker students did not ensure they had learned the topics independently.

- Kappa and Xi addressed this contradiction by offering students the opportunity to engage in collaboration formally in a group project. In these projects, the ways in which the students collaborated were more closely scrutinised. In effect, these projects created the opportunity for students to learn how to collaborate effectively when dealing with newly acquired mathematical knowledge and unfamiliar contexts.
<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Interpretation</th>
</tr>
</thead>
</table>
| Kappa    | • Encouraging of collaboration  
|          | • Concerned about cheating; therefore invigilated the summative test |
| Lambda   | • Encouraging of collaboration  
|          | • Was concerned he might appear less approachable than students  
|          | • Concerned about cheating; therefore invigilated the summative test |
| Mu       | • Encouraging of collaboration  
|          | • Wished for students to discuss the mathematics between themselves |
|          | • Was not concerned about cheating |
| Nu       | • Encouraging of collaboration  
|          | • Believed that students may be cheating inadvertently  
|          | • Was not concerned about cheating |
| Xi       | • Encouraging of collaboration  
|          | • Concerned about students being isolated  
|          | • Concerned about cheating; therefore invigilated the summative test |
| Omicron  | • Encouraging of collaboration  
|          | • Concerned about students not collaborating effectively |
|          | • Was not concerned about cheating |

**Table 5.11:** Lecturers’ interpretations of collaboration in CAA testing
Mu and Omicron believed that the students should develop the autonomy to learn how to work effectively, whether alone or with others. They suggested that those students that were collaborating ineffectively and were relying on others to aid them in assessments would be caught out in the exams anyway.

Lambda was concerned that students may have worked too independently and would rather students have collaborated if they viewed him as unapproachable or daunting.

Although Nu had expressed concern that students might have cheated in some tests and that there were certain actions that he would have construed as cheating, he held a similar view to that of Mu and Omicron: the effects of ineffective or inappropriate collaboration were minimal in the long term.

Establishing the extent to which the CAA system proved effective for the lecturers demands the examination of the lecturers’ aims for this kind of assessment and whether those aims were satisfied.

The interviews provided evidence that the lecturers had adopted many aims for CAA: some of these were achieved and others were not. Those others that were not achieved caused some of the contradictions that the lecturers faced.

Table 5.12 shows that there was one lecturer that was content that their aims were satisfied by using the CAA system and had adopted practices that allowed them to overcome the contradictions they experienced. In Lambda’s case, his use of the CAA system had reached a stable practice.

For the other lecturers, contradictions remain because there are aims that are not satisfied by the CAA system. Therefore, it is possible that the practices that these lecturers employed had not reached a steady state and there may have been changes to those practices since.

However, it is important to note that many lecturers had already responded to the fact that the CAA system did not satisfy those aims. Kappa, Mu and Xi stated in their interviews that other types of assessment were in place to satisfy their remaining aims. In the cases of Mu and Xi, there was regret that the CAA system was not able to satisfy all their aims for assessment. However, Kappa was content that her aims for assessment were satisfied by offering a range of assessment types. She felt that only offering assessment through the CAA system would leave mathematics learning “impoverished” and, therefore, offering projects to her students resulted in her use of the CAA system being more effective.
<table>
<thead>
<tr>
<th>Lecturer</th>
<th>Aims satisfied</th>
<th>Aims not satisfied</th>
</tr>
</thead>
</table>
| Kappa    | • Enable students to practise  
          | • Test students quickly and save time | • Test conceptual understanding |
| Lambda   | • Encourage students to work between lectures  
          | • Enable students to practise  
          | • Differentiate between students  
          | • Prepare students for a variety of contexts | |
| Mu       | • Test a large number of students quickly and efficiently  
          | • Test students on their ability to carry out procedures  
          | • Enable students to practise | • Establish the extent to which the students have grasped the content  
          | | • Test whether students are capable beyond the lecture content |
| Nu       | • Encourage students to work between lectures  
          | • Consolidate learning that happens in lectures | • Expand and develop students’ views of what “doing mathematics” is |
| Xi       | • Provide a fair means of assessment  
          | • Enable students to practise | • Encourage students that are struggling with the content |
| Omicron  | • Allow students to monitor their progress  
          | • Enable students to practise  
          | • Provide a convenient form of assessment for students | • Test conceptual understanding |

Table 5.12: Lecturers’ aims that were satisfied and not satisfied by the CAA system
This leads to the conclusion that Lambda found the CAA system to be most effective, albeit having replaced the summative test offered by the CAA system with his own paper test. Doing so allowed him to resolve his own contradictions when using CAA: he wanted to retain a personal connection to his students during assessment and he wanted to expand his students’ experiences of the contexts in which the mathematics they were learning could be used. Although this meant that more time was required to mark the summative assessment — which, in itself, created a contradiction against the promise that CAA would save time in assessment — he viewed this compromise as necessary and reasonable.

The other lecturers found the CAA system to be an effective assessment to an extent; however, the scope of effectiveness appeared to be less than for Lambda. Kappa had reached a stable practice for her use of the CAA system and found the system to be effective for her goals. However, the effectiveness of the CAA system relied upon the use of an additional means of assessment to satisfy the goals that could not be reached with CAA alone. Therefore, Kappa had reached a compromise in which she forewent one of her aims for assessment when evaluating CAA on the satisfaction that it would be accomplished by the group projects.

For Mu and Nu, CAA practice also appeared to be stable. The main concern was to encourage students to practise between lectures and to become accomplished in the carrying out of the necessary procedures that were taught in the lectures. They conceded that the CAA system was not effective for testing students beyond lecture content or for testing conceptual understanding. Mu and Nu believed that the low contribution to the module score from CAA scores is reflected by this lack of capability. Consequently, they believed that the CAA system offered an important aspect to the course, but other assessment types were required to fulfil their aims for assessment.

There remained concerns that the CAA system was not satisfying aims that were fundamental to Xi’s requirements for assessment. She was particularly concerned that struggling students were not able to progress under this assessment regime. While it satisfied her other aims, Xi was unable to resolve this contradiction at the time of the study. Therefore, the author concludes that her practice was prone to change as she sought to overcome this contradiction.

Xi was particularly attuned to her students’ needs and concerns. Indeed, it was this awareness that made her more knowledgeable of her students’ actions during CAA testing and the shortcomings of the CAA system that raised this contradiction. This made the likelihood of a change in practice greater.
Omicron was keen to change the CAA system in its entirety. Although Xi appeared to be the least satisfied with the system, Omicron was the only lecturer that had expressed a desire to change the system and had made enquiries to make such a change. Although her practice and outlook with regard to the CAA system compared with Mu and Nu, Omicron gave the clearest indication that the lack of the capability to test conceptual understanding presented a significant contradiction for her practice. Nonetheless, Omicron remained content that the existing system satisfied some of her aims for assessment: it enabled students to practise between lectures and it was convenient for them. However, she believed that other CAA systems would satisfy these aims equally well and could offer the capability to assess conceptual understanding. Thus while the CAA system was effective for Omicron’s aims to an extent, it was not effective for the most important of her aims.

5.2.3 Summary

Lecturers are intrinsic to the activity of CAA. Ultimately, they determine how the tests are administered, which questions appear in the tests, and to what extent the tests have a role in determining the students’ overall achievement in their modules. While students have their own aims for assessment, it is clear that lecturers have their own aims and how they use the CAA system is, in part, determined by what they wish to achieve.

All of the lecturers were subject to the expectations of the mathematics department — and for those teaching mathematics to students from the engineering departments, additional constraints may have applied. Yet above and beyond these requirements, this study identified a number of different motives that lecturers have for assessment more generally.

Some of the lecturers saw summative assessment as preparation for the module exam and for differentiating between students of different levels of accomplishment. Some regarded summative assessment as an opportunity to provide feedback to the student and encourage them to practise between lectures. Some felt that summative assessment was the opportunity to assess students beyond the lecture content and to determine the extent to which students understood concepts.

These requirements that are not imposed onto them by the department and the wider university community may have come from their identities as lecturers, mathematicians, engineers and researchers. These identities sometimes cause primary contradictions Engeström (2014), which results in one identity becoming a dominant, leading identity (Black et al., 2009).
The leading identity for an activity develops within a figured world (Holland et al., 1998a). Figured worlds are imaginary spaces that individuals inhabit, making inferences about the values of their communities. For a lecturer, this space includes the culture and community of a university; their interactions with students, other lecturers and researchers; and their histories. Gustin (1985) noted that lecturers formed their own notions about mathematics during their studies; in particular, they formed their own identities as mathematicians as students and adopted elements of their teachers’ approaches to mathematics. The complexities of the formation of their leading identity resulted in these individual approaches to assessment and differing opinions on what is important.

Consequently, lecturers’ practices differed as a result of their need to satisfy their requirements and those of their departments. This resulted not only in differences between their CAA practices, but also in their approach to assessment regimes in the modules overall.

The lecturers found that the CAA system was effective for the requirements of the department: it was an apparently ‘fair’ assessment in that the students were marked consistently and the students obtained the same level of feedback. It also largely satisfied the needs that they had in common: it encouraged students to practise between lectures and they received timely feedback.

However, they found that there were some aims that were not satisfied by the CAA system. The system was not able to assess students on their awareness of concepts to the extent that some of the lecturers required. A few lecturers commented that students were not being exposed to different contexts or being tested beyond the lecture content.

In response to these contradictions, some of the lecturers have been reviewing their practice. Some lecturers have used paper tests instead of invigilated online tests to add rigour and diverse contexts. Some lecturers added a coursework component to their assessment regime in order to assess their students’ conceptual understanding.

This led to a diversity of practice and some residual dissatisfaction with the CAA system. While the lecturers had reached settled states in terms of their practice, there remained a desire from some lecturers to change the system in its entirety. For these lecturers, the CAA system fulfilled an important function in allowing students access to regular practice and feedback; however, they believed there was potential in a new system to address the shortcomings and contradictions that remained.
5.3 Summary

The lecturers and students that participated in these studies — and the interviews, in particular — gave the opportunity to critique the CAA system for its effectiveness and where there were opportunities for improvement. This summary uses the findings from the lecturers and the students to answer the three research questions.

5.3.1 What makes an assessment “effective”?

The interview studies confirmed that each individual has their own aims for assessment, which are not merely due to their role in assessment. Between students, there are those that are aiming to get full marks in assessment, and those that are merely wishing to pass. Nonetheless, all the students related their aims in terms of the marks that were on offer. They could appreciate the cultural value of earning marks, since they would inevitably lead to their module score and, eventually, their degree classification.

Some students had other motives for learning. They might have had particular vocations in mind when choosing their subject of study, and their work towards a degree was motivated by a desire to get a particular job. In this case, the students’ identities informed their goals. In some cases, this meant that they were seeking knowledge and understanding that would be useful in that role. For them, an effective assessment would ideally have helped them to acquire that knowledge and understanding.

One student in the study was motivated by deepening his understanding for its own sake. He regarded mathematics as intriguing and interesting. For this kind of student, an assessment tool that would test conceptual understanding and give challenge beyond the expectations for procedural awareness would be more effective.

There were also differences in what lecturers were hoping to achieve in assessment. Most were keen to encourage students to practise using the procedures they had been taught in lectures, but the awarding of marks was a low priority. However, the lecturers were pressed for time since they had other responsibilities. Thus assessment exercises that were quick and straightforward to administer to students and handle feedback would be effective for lecturers.

However, they also wished to assess a wide range of skills and understanding. In terms of mathematics assessment, this can be difficult since mathematics itself can be difficult to communicate. There was also a desire to encourage students to become more autonomous agents in assessment: ideally students would develop the skills to monitor
their performance and identify their own weaknesses through self-regulation and assessment.

There was also a desire for assessment to be objectively fair and consistent, which was also a requirement from the department and the university. Students also desired this fairness and consistency; they wished to be credited fairly for what they knew and understood.

5.3.2 Under what conditions is the CAA system an effective assessment?

The CAA system was effective for all the students in the interview study to some extent. All the students in this study were aiming to achieve a particular mark in the assessment and, for the most part, the practice tests enabled them to achieve these aims. This made the practice tests an intrinsic component of the assessment; from the interview data, it would seem that the CAA system might not have been as effective had it consisted solely of the summative tests.

However, the scope for the system to be effective was limited by the students’ inability to choose more challenging learning goals after the summative test. The summative test was regarded as the end of learning, after which they could gain no further credit for learning.

As for the students, the CAA system was effective for the lecturers in the interview study to some extent. It offered regular assessments to students, who could take responsibility for managing their own time to complete the assessments, and gain immediate feedback from which they could improve.

It also achieved the lecturers’ aims to encourage students to use the lecture content on problems between the lectures. Indeed, the students reported in the questionnaires and in the interviews that they would commit substantial time and effort to ensuring that they could carry out the procedures that were asked of them for the summative test.

The tests were also regarded by students and lecturers as being objectively fair. While randomisation of the questions minimised the possibility that two students would face exactly the same test, the questions were similar enough between tests to ensure there were no differences in difficulty. Students also trusted the CAA system to be accurate in its marking, though the lecturers had reported instances when the solutions stored in the CAA system were incorrect.

Both lecturers and students were broadly satisfied with the feedback, insomuch that the students could identify where they had gone wrong and improved their approach to
the question. However, it was a concern for the lecturers that the students were too dependent on the feedback, and some had learned to mimic the methods given in the solutions. It was also reported in an earlier focus group that some students believed that it was possible to pass the test without attending the lectures (Broughton et al., 2012).

5.3.3 What would make the CAA system more effective as an assessment tool?

Identifying where the assessment ceases to be effective offers the point at which the assessment could be made more effective. For the students, the limitations come largely after the summative assessment. This occurs because students are unwilling or unable to acquire new learning goals.

The lecturers also feel that the CAA summative test results are not necessarily good measures of a student’s understanding or learning. That is, the goals that students set for themselves in terms of marks do not reflect the depth of understanding that lecturers desire to see in their students.

These indicate that there is a need for students to acquire learning goals after the CAA summative test, or for the CAA system to be able to assess students to a depth of understanding that is not currently available.

In Eta’s case, he had goals beyond what the CAA tests could offer him. He desired more conceptual understanding and wished the CAA tests could guide him towards his goals. It is not just the CAA system that prohibited Eta to achieve his learning goals; he also found that his other learning commitments and the lack of time made pursuing his goals impossible. Consequently, the effectiveness of the CAA system ceased once he had completed the summative test.

Both lecturers and students found that time was a barrier towards extending the scope of effectiveness for the CAA system. The students were under pressure to complete assignments and were managing their time between different modules. Lecturers found it hard to find time to make improvements to the CAA system: Nu had attempted to write more questions, but found this process was not straightforward and yielded little benefit.

The lecturers tended to use the CAA system as an exercise for students to complete to supplement their learning. The marks they were allocated were sufficient enough to encourage students to use the tests, but not substantial enough to make a significant contribution to their overall module grades. For the mathematics lecturers, this meant that cheating was not unduly concerning; for the engineering lecturers, there remained
a desire for the marks to indicate where students were struggling. The engineering lecturers were keen to prevent students from collaborating the summative test, and so used invigilation of the summative tests to prevent cheating.

In order to address these contradictions between their aims and what the CAA system offered, the lecturers adopted diverse practices. Some used alternative forms of assessment to assess students on their conceptual understanding; one lecturer used summative tests conducted on paper instead of online to introduce real-life contexts and additional challenge to students.

While these diverse practices extended the effectiveness of the lecturers’ assessment regime, this diversity had an impact on the effectiveness of the CAA system for students. Theta found that the change in practice affected her ability to pass the tests. It would seem that Theta had become dependent on the context of the CAA system questions, which resulted in her becoming confused when real-life contexts were introduced in the second semester.

Theta’s scenario showed that there was a problem with the contextual dependency on the CAA system, and that some students could not apply their knowledge to a context that was not familiar to them. For the lecturers that conducted the summative test in this way, it helped to identify these shortcomings in the students’ awareness and to differentiate between students’ levels of understanding. However, Theta merely wished to pass the test and neglected to return to the practice tests after failing the summative test; the CAA system ceased being effective for her at this point.

This case, too, exemplifies the belief that a summative test represents the conclusion of learning. The students rarely revisited the practice tests after the summative test, though some did so to revise for their exams. The lack of a tangible reward for continuing learning beyond the summative test is pervasive and was a significant limiting factor to the scope of effectiveness for the CAA system.

It highlights a need for a shift in the community towards a cultural reward for continued study. Although the marks awarded cannot change after a summative test, there may be other rewards to be had — for example, it may be encouraged as a means to prepare for a future course, or for a side project, or for revision for exams. The interview study with students suggested that there needs to be an ulterior motive for students to invest time in further study: even when students are already engaged and interested in pursuing deeper understanding.

This could also be achieved by ensuring that the CAA system offers the depth of challenge that lecturers desire. However, while Nu had attempted to write new questions for the system, he did not find it to be worthwhile and required significant time investment. The
lecturers, as a whole, were not convinced that the CAA system could assess students to level of conceptual understanding that they desire. This led to some lecturers using an alternative assessment tool, used alongside the CAA tests, and some to replace aspects of the CAA system with a paper analogue in which the questions could probe a student’s conceptual understanding.

While these responses from the lecturers mitigated against the limiting factors of the CAA system’s scope of effectiveness, they do not address the limiting factors themselves. The students’ reticence towards setting new goals — particularly towards those aimed at their conceptual understanding — could be addressed by setting questions that aim to test their understanding to this level. Nu had tried to achieve this, and found that the process was time-consuming: investment in developing questions that tackle conceptual understanding could address this problem.

Lambda had replaced the summative CAA test with a paper test, since he felt that students were becoming dependent upon the context of the CAA questions. This was observed in another module, where Theta failed tests in her second semester in a similar paper-based test. This somewhat legitimises Lambda’s concerns about students’ context-dependency. This could be addressed by introducing questions that test students’ ability to apply their knowledge to unfamiliar contexts — Lambda was keen to introduce feasible, real-world problems for students to solve using their prior knowledge.
Chapter 6

Conclusions

The purpose of this final chapter is to review the findings of the student and lecturer studies in light of what this project intended to achieve. It begins with a discussion of the initial problem in evaluating an existing assessment tool and how the model for effective assessment was derived from the literature and the theoretical framework. Next, the key findings from the studies address the research questions and provide a response to the initial problem. Finally, this chapter discusses how the findings of this research may be used elsewhere.

6.1 How can an assessment tool be evaluated?

The purpose of this research was to evaluate a computer-aided assessment (CAA) system. However, the literature exposed the lack of a rigorous evaluation of mathematics-based computer-aided assessment: either for specific systems or for its use generally. There are many practitioner-based accounts of introducing such systems in institutions or individual courses; many of these reported the positive effects and lacked a critical evaluation, leaving many long-term effects unreported.

The intention of this study was to examine the scope of effectiveness of one CAA system for both student and lecturer users and to provide such an evaluation. An earlier focus group study and informal conversations with lecturers had offered some insights into their practices and some of the issues they had experienced.

Choosing a theoretical framework was important for developing an evaluation framework informed by theory and literature. Cultural-historical activity theory (CHAT) offered a framework that modelled individuals engaged in social activity, motivated by need and mediated by tools. Assessment is a typical human activity in that regard, and assessment
is a culturally-created tool that enables students to prove their learning. Thus the ability to demonstrate that learning has taken place is crucial to the success of an assessment; calling it “effective assessment” defined this characteristic of the learning tool.

Cultural-historical activity informed several important elements of the model for effective assessment. CHAT holds that all human activity is purposeful and that individuals are influenced by their culture and history when forming goals. Therefore, students and lecturers will always have aims when embarking on assessment, and they will form their aims individually with influences from colleagues and further afield. This model offers a measure of an assessment’s effectiveness; a critical narrative of students’ and lecturers’ use of the assessment tool identifies where the assessment has been effective and is no longer effective for those individuals.

It was important to establish the students’ and lecturers’ aims in order to understand why their activity diversified. The model for effective assessment proposed that effectiveness should be measured with respect to the learning aims of students and lecturers; in other words, an assessment would be deemed to be effective while it fulfils the intentions of those using it.

The notion of internalising and externalising information in the learning process arose from Engeström’s (2001) work on expansive learning: students acquire knowledge from external sources and internalise this information; they must then externalise this information to demonstrate their learning. Therefore, the model for effective assessment holds that assessment is an essential part of the learning process. Assessment defines the opportunity to demonstrate that learning has taken place. Without assessment, there is no evidence that learning has happened.

The assessment literature suggested that the process of learning ought to encourage students to become less dependent and more autonomous. CHAT also argues that goal-forming is cyclic: Both Engeström (2001) and Tikhomirov (1988) described goal formation as cyclic: once a goal is reached, the individual should become more aware of what else is achievable. Therefore, it is a condition of the model for effective assessment that, in order for an assessment to remain effective, students must develop autonomy and pursue more challenging goals.

While CHAT emphasises the social nature of activity, the model for effective assessment emphasises that individuals are unique insomuch that their histories and identities are unique, and therefore it follows that their aims and actions are unique. So while an assessment tool may be effective for one lecturer, it might not be effective for another lecturer or for the students. However, deductions can be made from looking at several
case studies and identifying where an assessment has been effective and common causes for it ceasing to be effective.

6.2 Is the CAA system an effective assessment?

When developing a model for effective assessment, the natural question to ask is whether an assessment tool is effective; however, this is the wrong question. No assessment tool could ever be always effective and in every case. The answer will usually be “yes, and no”.

The model for effective assessment appreciates that students and lecturers are individuals that seek different things from learning and, therefore, from assessment. Since their aims are different, the assessment tool that is presented to them may differ in its effectiveness. In other words, the extent to which an assessment is effective depends on individuals’ goals and their identities.

There are also societal and cultural impacts on individuals’ identities and their goals. Goals may be influenced by what is culturally significant and rewarding in society. For students, marks were commonly referenced as goals for the CAA tests. Students wish to acquire marks because they have cultural value and significance: they will contribute towards their degree classifications which, in turn, affects the student’s future job prospects. Once the opportunity to achieve — or fail to achieve — those goals have passed, the students did not have the motivation to pursue more challenging goals. This may be because there were no tangible cultural rewards for doing so: there were no further marks to be acquired, so there was an implicit cultural message that further study is not as valued.

With respect to the lecturers that were using the CAA system, there is an analogous endeavour towards reaching a relatively stable practice where there are no contradictions between practice and what the lecturer aims to achieve in assessment. Engeström argued that instability and contradictions are the cause of further change — contradictions generate “disturbances and conflicts, but also innovative attempts to change the activity” (Engeström, 2001, p. 134) — and that this change affects others in the environment too (Daniels, 2004).

One can envisage how a change to assessment practice during a course could have an impact on students: it may create confusion, for instance. However, even if the change occurs between cohorts, then there may be an effect whereby students that have previously studied on the course cannot give effective advice to members of the new cohort about the assessment. Nonetheless, the position that CHAT holds is that change occurs as a product of contradiction. Therefore, change is an attempt to resolve a problem and,
thus, bring an improvement. It must be noted with caution, however, since these changes have an impact on the community that may not always be positive.

The state of reaching a stable practice does not indicate that there are no contradictions: they “can either enable learning to progress, or they can actually disable it, depending on whether or not they are acknowledged or resolved” (Murphy & Rodríguez-Manzanares, 2008, p. 445). Thus there may be unknown contradictions, or issues that are simply not acted upon. From an analysis perspective, they are difficult to identify and discuss. However, it does make it important to note that an apparently stable practice does not necessarily mean it is a perfect practice. Hence, a stable practice does not necessarily mean the practice is effective.

The lecturers in this study had each reached practices that were steady. Some lecturers were interested in changing the CAA system, or viewing other ways of incorporating tests of conceptual understanding into the system, so some contradictions remained. The presence of contradictions motivate change and, consequently, the CAA system was replaced for mathematics students after the lecturer study had been completed.

### 6.3 How can this model for effective assessment be used elsewhere?

A potential criticism of the model is that any evaluation of an assessment tool will have broadly the same conclusion: it will be effective sometimes, and in some circumstances, but not always.

However, through this research and by using the model, points were identified at which the assessment ceased to be effective. From these points, potential changes to the assessment tool could have extended the scope of its effectiveness.

The lecturers had already made adjustments to their practice in order to extend the scope of the effectiveness of the CAA system. This had an impact on the students, since they encountered different practices themselves and sometimes found it difficult to adjust. Thus while lecturers yielded greater effectiveness from the assessment tool, this came at the cost of the scope of effectiveness with respect to some students.

This study highlighted the problem that students were somewhat dependent on marks for motivation to study. While some students may have been able to make further use of the practice tests to ensure they could carry out the procedures, they often did not. The summative aspect of the test is culturally understood to be the final score and there is no cultural reward after it. To that end, this research suggests there is a need to change
the learning culture towards an intrinsic reward for working beyond the summative test. Assessment could be designed or modified to ensure that these higher goals are tested.

It is the intention that other assessment tools may be evaluated using this model for effective assessment. Not only did it offer a critique of the current use from different perspectives, it also offered opportunities to make adjustments that may extend the scope of effectiveness.
Appendix A

CAA system examples

All figures in Appendix A are from the HELM Project (2006).

Figure A.1: An example of one of the questions in the CAA system.
Appendix 159

Figure A.2: An example of a results page in the CAA system

Solution

Consider

$$\frac{Mx + N}{(x^2 + \beta)(x - \alpha)} = \frac{Ax + B}{(x^2 + \beta)} + \frac{C}{(x - \alpha)}$$

Simplify the expression on the right-hand side by finding the common denominator. That is,

$$\frac{Mx + N}{(x^2 + \beta)(x - \alpha)} = \frac{(Ax + B)(x - \alpha) + C(x^2 + \beta)}{(x^2 + \beta)(x - \alpha)}$$

Now simply equate powers of $x$ in the numerators on each side:

$$0 = A + C, \quad M = -A\alpha + B \quad \text{and} \quad N = C\beta - B\alpha$$

Solve these three equations to find $A$, $B$ and $C$.

Figure A.3: An example of the feedback given by the CAA system
Appendix B

Student questionnaires

B.1 Semester 1 questionnaire
Dear student,

My name is Stephen Broughton. This questionnaire is part of my PhD project that examines the effectiveness of feedback in computer-aided assessment (CAA). It is likely you will experience some practice tests or assessed tests using a computerised system in mathematics modules. This questionnaire is designed to help us understand what new students expect from CAA.

The information you provide will be confidential. We ask that you provide your student number so that we can match your responses later in the year.

We are very grateful for your participation and helping us in our ultimate aim of improving CAA.

Your student number

Relationships with Mathematics
1. In my future studies, mathematics will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

2. Ensuring I understand all the mathematics in my future studies will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

3. If my future studies feature more mathematics than I had expected, I would be
   - very happy
   - fairly happy
   - neither happy nor unhappy
   - fairly unhappy
   - very unhappy

4. In my future studies, I would prefer there to be
   - a lot of mathematics
   - quite a lot of mathematics
   - some mathematics
   - as little mathematics as possible
   - no mathematics
   - not sure

5. When studying mathematics in my future studies, I hope it will be
   - familiar mathematics that I already know
   - new mathematics that I have not done before
   - a mixture of familiar and new mathematics

6. My future career will be in
   - financial
   - engineering
   - teaching
   - other (please state) ________________
   - not sure

7. In my future career, I think mathematics will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

8. What degree classification are you aiming for? (Please tick one)
   - First class  [ ] (90-100 marks)
   - Second class Division I [ ] (80-89 marks)
   - Second class Division II [ ] (70-79 marks)
   - Third [ ] (60-69 marks)
   - Pass [ ] (50-59 marks)
   - Third [ ] (40-49 marks)
   - Pass [ ] (35-39 marks)
   - Have not considered my degree classification [ ]

Office use

COMPUTER-AIDED ASSESSMENT SURVEY

Your student number

Relationships with Mathematics
1. In my future studies, mathematics will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

2. Ensuring I understand all the mathematics in my future studies will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

3. If my future studies feature more mathematics than I had expected, I would be
   - very happy
   - fairly happy
   - neither happy nor unhappy
   - fairly unhappy
   - very unhappy

4. In my future studies, I would prefer there to be
   - a lot of mathematics
   - quite a lot of mathematics
   - some mathematics
   - as little mathematics as possible
   - no mathematics
   - not sure

5. When studying mathematics in my future studies, I hope it will be
   - familiar mathematics that I already know
   - new mathematics that I have not done before
   - a mixture of familiar and new mathematics

6. My future career will be in
   - financial
   - engineering
   - teaching
   - other (please state) ________________
   - not sure

7. In my future career, I think mathematics will be
   - essential
   - very important
   - not very important
   - not important at all
   - not sure

8. What degree classification are you aiming for? (Please tick one)
   - First class  [ ] (90-100 marks)
   - Second class Division I [ ] (80-89 marks)
   - Second class Division II [ ] (70-79 marks)
   - Third [ ] (60-69 marks)
   - Pass [ ] (50-59 marks)
   - Third [ ] (40-49 marks)
   - Pass [ ] (35-39 marks)
   - Have not considered my degree classification [ ]
Approach to Assessment

9 In mathematics homework, assignments and tests, which of these resources have you used for help? (Please tick all those that apply)
- [ ] Class/lecture notes
- [ ] Friends
- [ ] Revision books
- [ ] Teacher/lecturer
- [ ] Internet resources
- [ ] Other (please state)

10 I use these resources (please tick all those that apply)
- [ ] when I'm stuck
- [ ] before I start, to prepare
- [ ] at the end, to check

11 How strongly do you agree with the following statements?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Slightly disagree</th>
<th>Neither agree nor disagree</th>
<th>Slightly agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I prefer to work on problems in a group.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When my work is marked, I check where I have made mistakes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I find practising lots of similar questions helps me to understand.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12 Which of these words would you associate with ‘assessment in mathematics’? (Please tick all that apply)

- [ ] Easy
- [ ] Pressure
- [ ] Risky
- [ ] Quick
- [ ] Restrictive
- [ ] Lazy
- [ ] Revision
- [ ] Enjoyable
- [ ] Other (please state)
- [ ] Safe
- [ ] Challenging
- [ ] Practice
- [ ] Time-consuming
- [ ] Boring
- [ ] Perseverance
- [ ] Personal
- [ ] Impersonal
- [ ] Quick
- [ ] Enjoyable
- [ ] Other (please state)

13 Which of these words would you associate with ‘computer-aided assessment’? (Please tick all that apply)

- [ ] Easy
- [ ] Pressure
- [ ] Risky
- [ ] Quick
- [ ] Restrictive
- [ ] Lazy
- [ ] Revision
- [ ] Enjoyable
- [ ] Other (please state)
- [ ] Safe
- [ ] Challenging
- [ ] Practice
- [ ] Time-consuming
- [ ] Boring
- [ ] Perseverance
- [ ] Personal
- [ ] Impersonal

14 Are there any additional comments that you wish to make about mathematics, assessment or computer-aided assessment? Are there any comments that you wish to make about the questionnaire?

15 Would you be willing to take part in an interview to discuss CAA later in the year?
- [ ] Yes
- [ ] No

If you ticked yes, please give us two ways in which we can contact you in the future (e.g. email addresses, mobile number, etc.).

First contact preference ____________________ Second contact preference ____________________

You have the right to withdraw at any stage.
B.2   Semester 2 questionnaire
**COMPUTER-AIDED ASSESSMENT SURVEY**

Dear student,

My name is Stephen Broughton. This second questionnaire is designed to help us understand how students use computer-aided assessment (CAA) and how students feel towards mathematics assessment.

The information you provide will be confidential. We ask that you provide your student number so that we can match your responses from earlier in the year.

We are very grateful for your participation and helping us in our ultimate aim of improving CAA.

Your student number

### Section 1: Practice tests

1. **How many times do you usually attempt mathematics practice tests? (Please tick one.)**
   - 0
   - 1-2
   - 3-4
   - 5 or more
   - not sure

2. **When working on practice tests in this module... (Please tick one.)**
   - collaboration is encouraged
   - collaboration is discouraged/prohibited
   - not sure

### Section 2: What resources do you use to help you in the practice tests? (Please tick all those that apply.)

- Class/lecture/tutorial notes
- Lecturers/tutorial leaders/personal tutors
- Mathematics Learning Support Centre
- Yellow formulae book
- Textbooks
- Calculator
- Internet resources
- Mathematical software (e.g. Maple)
- Other (please state)
- I don’t use any resources
- Not applicable

### Section 3: In my mathematics modules...

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Slightly disagree</th>
<th>Neither agree nor disagree</th>
<th>Slightly agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I use practice tests to get better scores in the coursework test.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I keep using practice tests until I get the score I want.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice tests help me to identify my weak areas.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I prefer to work on practice tests in a pair or a group.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When practice tests are marked, I check where I have made mistakes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am happy with the support available from lecturers for practice tests.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I find practising using practice tests helps me to understand.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Section 2: Coursework tests (tests that are either CAA or on paper)**

5. **What resources do you use during coursework tests? (Please tick all those that apply.)**
   - Class/lecture/tutorial notes
   - Lecturers/tutorial leaders/personal tutors
   - Mathematics Learning Support Centre
   - Yellow formulae book
   - Textbooks
   - Calculator
   - Internet resources
   - Mathematical software (e.g. Maple)
   - Other (please state) _________________
   - I don’t use any resources
   - Not applicable

6. **What kinds of maths coursework tests have you experienced? (Please tick all those that apply.)**
   - In-class paper test
   - In-class computer test
   - Computer test taken anywhere
   - Other (please state) _________________

7. **When working on maths coursework tests in this module... (Please tick one.)**
   - collaboration is encouraged
   - collaboration is discouraged/prohibited
   - not sure
   - not applicable since test is invigilated

8. **What score do you aim for in your maths coursework tests? (Please tick one.)**
   - 90-100%
   - 80-89%
   - 70-79%
   - 60-69%
   - 50-59%
   - A score less than 50%
   - I don’t aim for a score

9. **In this module...**
   - It is easy to get 100% in the coursework test.
   - I would prefer to work on coursework tests in a pair or in a group.

10. **The coursework test is good at testing...**
    - my ability to recall mathematical facts, rules and equations.
    - my ability to carry out mathematical procedures and methods.
    - my deeper understanding of mathematical concepts.

11. **Which of these words would you associate with 'computer-aided assessment'? (Please tick all those that apply.)**
    - Easy
    - Pressure
    - Risky
    - Quick
    - Scary
    - Restrictive
    - Lazy
    - Revision
    - Enjoyable
    - Other (please state)

    - Safe
    - Challenging
    - Practice
    - Time-consuming
    - Perseverance
    - Personal
    - Impersonal
    - Boring
# Section 3: Your aims and beliefs

## 12 What degree classification are you aiming for? (Please tick one.)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>First class</td>
<td>(90-100 marks)</td>
</tr>
<tr>
<td>Second class Division I</td>
<td>(80-89 marks)</td>
</tr>
<tr>
<td>Second class Division II</td>
<td>(70-79 marks)</td>
</tr>
<tr>
<td>Third</td>
<td>(60-69 marks)</td>
</tr>
<tr>
<td>Pass</td>
<td>(50-59 marks)</td>
</tr>
<tr>
<td></td>
<td>(40-49 marks)</td>
</tr>
<tr>
<td></td>
<td>(35-39 marks)</td>
</tr>
</tbody>
</table>

- I have not considered my degree classification

## 13 How strongly do you agree with the following statements?

**Coursework** should help to identify my weak areas.

**Coursework** should help to build my confidence.

**Coursework** should challenge me.

**Coursework** should help me to improve.

**Coursework** should have easy and hard parts.

**Coursework** should test me beyond lecture content.

**Coursework** should give me detailed feedback.

## 14 How strongly do you agree with the following statements?

**Computer-aided assessment** helps to identify my weak areas.

**Computer-aided assessment** helps to build my confidence.

**Computer-aided assessment** challenges me.

**Computer-aided assessment** helps me to improve.

**Computer-aided assessment** has easy and hard parts.

**Computer-aided assessment** tests me beyond lecture content.

**Computer-aided assessment** gives me detailed feedback.

**Computer-aided assessment** is an effective coursework.

## 15 Are there any additional comments that you wish to make about mathematics computer-aided assessment?
Appendix C

Staff questionnaire
Dear ...,  
As you may be aware, my PhD project is examining the effectiveness of the feedback mechanisms in computer-aided assessment (CAA). This questionnaire will help us develop a profile of the procedures, issues and views lecturers have with CAA. We would be very grateful for your input to this project and we invite you to respond to these questions with regard to MAAXXX. Thank you, Stephen Broughton.

1. What aspects of CAA do you use with your students?  
☐ Practice tests  
☐ Coursework (summative) tests

2. What feedback do students receive from the CAA you use?  
☐ Scores for practice tests  
☐ Scores for coursework tests  
☐ Right/wrong indicators for practice test questions  
☐ Right/wrong indicators for coursework test questions  
☐ More detailed feedback from practice test questions  
☐ More detailed feedback from coursework test questions

3. When are practice tests available to students?  
☐ Up to one week before a coursework test  
☐ More than one week before a coursework test

4. Are practice tests accessible after coursework tests are completed?  
☐ Yes  
☐ No

5. Can students access practice tests while coursework tests are available?  
☐ Yes  
☐ No

6. How do students complete your coursework tests? (Please tick all that apply.)  
☐ Invigilated computer test in computer lab  
☐ Not invigilated computer test  
☐ Invigilated paper test that is similar to the CAA coursework test  
☐ Not invigilated paper test/coursework that is similar to the CAA coursework test  
☐ Other (please specify) __________________________

7. What is your policy on student collaboration in practice tests?  
Practice tests  
Coursework tests

☐ I encourage it, and I have communicated this to students.  
☐ I encourage it, but I have not communicated this to students.  
☐ I discourage it, but I have not communicated this to students.  
☐ I discourage it, and I have communicated this to students.

8. To what extent do you agree with the following statements?  

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAA tests students’ ability to recall mathematical facts, rules and equations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA tests students’ ability to carry out mathematical procedures and methods.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA tests students’ deeper understanding of mathematical concepts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I use CAA with my students because…

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>it is easy to set CAA tests for my students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>it was used by a previous lecturer.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am encouraged to by the department.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>it is departmental policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>students receive immediate feedback.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>students receive good quality feedback.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA frees up time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA is convenient.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA provides students with the opportunities to practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA provides students with the motivation to practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Have you developed or attempted to develop your own CAA questions?

- Yes
- No

To what extent do you agree with the following statements?

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing CAA questions takes too much time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It would be worthwhile to develop new questions each time this module is run.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is too difficult to develop new CAA questions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The existing CAA questions provide sufficient challenge for my students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is worth learning how to develop questions for the CAA system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are there any additional comments that you would like to make regarding CAA, its feedback or how it is used?

Would you be interested in taking part in an interview to discuss computer-aided assessment?

- Yes
- No
Appendix D

Example of questions used in the student interviews

• How has your first year been so far? Has it matched your expectations?

• What sort of student would you say you are? How well do you think you are doing?

• Tell me about your experiences of CAA so far this year.

• How would you prepare for a CAA test? Where do you do them? What else do you have with you when you do the practice tests? So you do anything different between the practice tests and the actual test?

• What do you aim for? Do you aim for particular scores when you’re doing CAA? How do your CAA marks compare with other assessments that you do? Do the marks motivate you to try harder on CAA? Would you do CAA if you receive no marks for it?

• How common do you think collaboration is in CAA? What are your feelings on collaboration? When does collaboration become cheating?

• Do you find CAA helpful? Does it help you to improve? What does it help you to improve on?

• What do you like about CAA? What don’t you like about CAA? Is CAA different from what you expected it to be at the beginning of the year?

• What makes an assessment effective for you? Is there anything that you wish to add?
Appendix E

Example of questions used in the lecturer interviews

- Why do you use CAA?
- Why is CAA set up this way in your module?
- What changes might you make in the near future?
- What does CAA test?
- How have your interactions with students and other lecturers influenced the way you use CAA?
- What are the reasons for your policy on collaboration between students in CAA activities?
- In what ways do students collaborate in CAA exercises?
Appendix F

Student questionnaire analysis

The preconceptions questionnaire (Q1) was completed by 688 students from across the eight disciplines; 369 students completed the experiences questionnaire (Q2). Of these students, 285 completed both. For most cohorts, the questionnaires were handed out in lectures or tutorials. In the second semester, there were fewer responses for two reasons: one, there were fewer students attending the lectures or tutorials that the questionnaires were distributed; and two, there was one cohort for which an electronic version of the questionnaire had to be distributed. The response rate for the electronic version was much lower.

F.1 Cohort attitudes towards mathematics

In the first questionnaire, the students were asked to complete the statement, “In my future studies, mathematics will be...”, by choosing from “essential”, “very important”, “not very important”, “not important at all”, or “not sure”. The responses across the eight

<table>
<thead>
<tr>
<th>Subject</th>
<th>Q1 only</th>
<th>Q2 only</th>
<th>Both Q1 and Q2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>102</td>
<td>18</td>
<td>90</td>
<td>210</td>
</tr>
<tr>
<td>Aero. and Auto. engineering</td>
<td>69</td>
<td>9</td>
<td>57</td>
<td>135</td>
</tr>
<tr>
<td>Civil engineering</td>
<td>43</td>
<td>13</td>
<td>37</td>
<td>93</td>
</tr>
<tr>
<td>Electrical engineering</td>
<td>43</td>
<td>3</td>
<td>3</td>
<td>49</td>
</tr>
<tr>
<td>Manufacturing engineering</td>
<td>16</td>
<td>14</td>
<td>31</td>
<td>61</td>
</tr>
<tr>
<td>Materials engineering</td>
<td>28</td>
<td>5</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td>Mechanical engineering</td>
<td>73</td>
<td>14</td>
<td>43</td>
<td>130</td>
</tr>
<tr>
<td>Sports technology</td>
<td>29</td>
<td>8</td>
<td>9</td>
<td>46</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>403</strong></td>
<td><strong>84</strong></td>
<td><strong>285</strong></td>
<td><strong>772</strong></td>
</tr>
</tbody>
</table>

Table F.1: The distribution of questionnaire responses across cohorts.
Appendix

In my future studies, mathematics will be…

<table>
<thead>
<tr>
<th>Disciplines</th>
<th>Essential</th>
<th>Very important</th>
<th>Not very important</th>
<th>Not important at all</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeronautical and automotive engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrical engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanical engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure F.1: Students’ perceptions of the importance of mathematics in their future studies, separated by cohort.

disciplines were quite similar: in every cohort, more than 80% of students believed that mathematics would be very important or essential to their studies (fig. F.1).

The students were also asked the extent to which mathematics would be important for their future careers (fig. F.2). In their responses, they suggested that most were interested in careers that demanded the use of mathematics; more than 70% of students in every cohort believed that mathematics would be very important or essential.

In another question, the students were asked their preference for the amount of mathematics they would be facing in their future studies (fig. F.3). The responses between the cohorts were more different: in particular, 80% of the sports technology wanted no more than “some mathematics” on their course; and no manufacturing students professed to wanting “a lot of mathematics”.

This opening analysis demonstrates that, across the disciplines, the students tended to accept that mathematics was both an important component of their degree courses and an inevitable part of their future careers. However, it also suggested that some groups were less confident or less willing to engage with mathematics during their studies.

The differences between the cohorts appears evident when observing some of the experiences that students reported in the second questionnaire (fig. F.4). For example, when asked how many times they usually attempted the practice tests, there were stark differences between the engineering cohorts in particular. In the Civil engineering group,
68\% of students usually took the test at least three times, whereas just over 27\% of the Mechanical engineering students took the test as many times.

It serves to be mindful that there may be some differences between students in different disciplines. There are some reasons for this: the mathematics entry requirement for the courses differ; and, as explained in the lecturers’ analysis, the delivery of CAA is different between the groups. It is also worth noting that the sample sizes for some cohorts were too small to perform meaningful statistical analyses exploring the differences between groups.

However, the intention of this project is to evaluate CAA using the definition for effective assessment; this requires determining whether CAA has been effective for individual students. It may transpire that the entry requirements or the implementation of CAA has an influence on the effectiveness of CAA for these students, but this must be determined from the individual students.

Further information from the questionnaire remains of interest, however the differences between cohorts are no longer sought. The students’ responses provide an overview of the entire first-year mathematics and engineering cohort that give some indication of the common approaches to completing CAA and issues that many students experience.
Appendix 175

In my future studies, I would prefer there to be…

![Figure F.3: The amount of mathematics that students are willing to engage in on their courses, separated by cohort.](image)

In both questionnaires, the students were asked to select words that they associated with CAA from a list of sixteen terms that were used by students in the focus groups. The terms and their context were not explained in the questionnaire, therefore it is not possible to give the reasons why students selected certain terms. Instead, the terms indicate some aspects of CAA that the students anticipated, in the case of the first questionnaire, or experienced, in the case of the second questionnaire (fig. F.5).

For these word associations, only the responses from those students that completed both questionnaires were considered. In the first questionnaire, 884 word associations were made (with a mean of 3.10 word selections per student); in the second questionnaire, 935 word associations were made (with mean 3.26 per student).

In the preconceptions questionnaire, the most popular words associated with CAA were Challenging (40.7\%\footnote{The popularity of words associated with CAA are expressed as the percentage of students that selected the word in the questionnaires}), Revision (32.6\%), Practice (32.3\%), Quick (29.1\%), Pressure...
Appendix 176

Figure F.4: The number of times students normally access the practice tests, separated by cohort.

(28.8%) and Restrictive (26.7%) — each of these words commanded more than 25% of the students making these associations.

In the experiences questionnaire, there were fewer words that were selected by more than 25% of the students, and these words also tended to have been selected by more than 40%. This is, perhaps, as expected since the students developed a much clearer impression of what CAA is and what it does. The most popular terms were Practice (48.4%), Quick (43.9%), Revision (40.7%) and Restrictive (28.4%).

Over the course of the first year, there appeared to be some shifts in the students’ feelings towards CAA: before attempting CAA, 28.8% of students associated CAA with pressure, but this had fallen to 13.7% of students towards the end of the year. Likewise, from being the most selected term in the first semester, the term ‘challenging’ fell from 40.7% to 22.1%. The terms that fell in popularity — pressure, challenging, risky, impersonal and time-consuming — could be argued to carry negative connotations; the fall in the number of students selecting those terms suggests that some students had overcome their cautious preconceptions of CAA.
Some of the terms attracted a broadly similar number of selections during the course of the year: there was a change of less than 3% for restrictive, safe, lazy, personal, impersonal, enjoyable and boring. Restrictive maintained selections of more than 25% of students, indicating that students anticipated that CAA would be restrictive in some way and experienced restrictions during the course of the year.

There were some terms that attracted much more of the students’ selections in the second semester than in the first: easy (10.5% to 18.3%), revision (32.6% to 40.7%), practice (32.3% to 48.4%) and quick (29.1% to 43.9%) commanded the greatest rises. These terms suggest that the students tended to find CAA to be more straightforward than anticipated: for many students, this related to the speed of getting through practice tests; for some, this related to the test material being quite straightforward.

The responses to the questions discussed to this point indicate that the students were, for the most part, confident, competent and prepared for the mathematics material they were anticipating during the course of the first year; however, they were unsure of what to expect for CAA and some students were anticipating CAA to be more daunting than it turned out to be. With over 40% of students selecting Practice and Quick, it
would appear that many students were prepared to make use of the practice tests and, perhaps, this is because they were seen to be quick. Far fewer students associated CAA with Challenging in the second semester than the first, which could be a concern.

Further questions in the questionnaire explored the approaches students took towards CAA, including the practice tests.

### F.2 The cohort’s approaches to CAA

In the preceding subsection, the students indicated that they were willing to undertake several practice tests to ready themselves for the summative test. The students were somewhat split when asked the extent to which they agree with the statement, “*It is easy to get 100% in the coursework test.*” Around 28.5% agreed or strongly agreed with the statement; 42.2% disagreed or strongly disagreed; the remaining 29.3% neither agreed nor disagreed (fig. F.6). It seems that quite a substantial minority of students found the coursework tests relatively straightforward to complete.

The students were also asked about their aims for degree classification — at the start of the year and at the end of the year — and their targets for coursework test scores. Figure F.7 shows that during the course of the first year, the students have maintained broadly similar ambitions to obtain first class degrees (77.3% at the start of the year compared to 81.0% at the end of the year). For the coursework tests, 89.3% of students aim for more than 70% of the marks available.

There were other questions in the experiences questionnaire that examined students’ approaches to CAA and give indications where these approaches might be different to other assessment types. For example, the students were asked how willing they were to work in groups to tackle problems in the first questionnaire; they were asked in the
second questionnaire the extent to which they preferred to work in groups in the practice tests. While 48.6% preferred to work in groups — at least at the start of the first year — they appeared less willing to do so in the CAA practice tests, with 27.8% indicating that they agree or strongly agree (fig. F.8).

In the experiences questionnaire, the students also gave indications of their goals for CAA. When presented with the phrase “I use practice tests to get better scores in the coursework test.”, 62.9% strongly agreed and 24.7% agreed — only 4.0% disagreed or strongly disagreed. For “I keep using practice tests until I get the score I want.”, 43.3% strongly agreed and 28.5% agreed, with 16.7% disagreeing or strongly disagreeing.
students also seemed to use practice tests to help to identify weak areas, with 49.1% strongly agreeing and 39.5% agreeing; 4.6% disagreed or strongly disagreed.

There is interest in the kind of knowledge and understanding that the coursework tests assess. With 82.3% indicating that either they strongly agreed or agreed with the statement, “I find practising using practice tests helps me to understand,” the natural question that arises is that of the depth to which this understanding is developed. The students were asked the extent to which they agree to the following statements.

- “The coursework test is good at testing my ability to recall mathematical facts, rules and equations.”

- “The coursework test is good at testing my ability to carry out mathematical procedures and methods.”

- “The coursework test is good at testing my deeper understanding of mathematical concepts.”

The students seemed to believe that the coursework test assessed recall and procedural skills (with 79.6% and 85.4%, respectively, that either agree or strongly agree), whereas 54.8% believed that it tests their conceptual understanding. Indeed, 20.2% of students disagreed, or strongly disagreed, that the coursework tests assessed their conceptual understanding, compared to 10.5% for recall and 8.1% for procedural skills (fig. F.9).
It would appear from this questionnaire item that the students felt that the coursework tests were stronger at assessing their procedural skills more strongly than the other skills. The questionnaire asked students to make comparisons between what they believed coursework should do and what they believed CAA achieves (fig. F.10).

![Comparison between what students believe an assessment should achieve and what students believe CAA does.](image)

**Figure F.10:** A comparison between what students believe an assessment should achieve and what students believe CAA does.

There were two aspects of coursework that nearly all students believed were required: it should help them to improve (95.2% agree or strongly agree), and it should give them detailed feedback (94.9%). However, over a quarter of the students felt that CAA did not give students detailed feedback: 26.8% either disagreed or strongly disagreed that it does. Although the students offered generally positive appraisals of CAA in the questionnaire, it would seem that its weakest aspects were offering detailed feedback, identifying weak areas (13.7% disagree or strongly disagree), building confidence (15.3%) and helping to improve (8.5%).

The students appeared to be uninterested in being tested beyond the lecture content — 42.9% of students either disagreed or strongly disagreed that coursework should test
beyond this content. Although 30.1% of students felt that CAA did not test beyond lecture content, this appears to be as these students expected.

In a similar manner, the questionnaire asked students whether they felt CAA was an effective assessment. For 63.5%, the students believed it was effective; for 16.6%, they felt it was not, with a further 19.9% that were seemingly unsure.

F.3 Discussion of the student questionnaires

The questionnaires were not designed to evaluate CAA according to the definition for effective assessment; rather, they offered some common views of and approaches to CAA and the coursework tests that follow.

When compared to the responses that students gave as requirements for coursework, it would appear that there are some students that believe that CAA does not provide as much assistance to their learning as an assessment should. For the 16.6% of students that believed that CAA was not effective, their other questionnaire responses identified some possible causes for their concerns: offering detailed feedback, identifying weak areas, building confidence and helping to improve.
Appendix G

Student interview case studies

G.1 Alpha

G.1.1 Identity and Aims

Alpha was a student from the Mathematics Department aiming for an upper second class degree classification (a 2:1). She wanted to embark on a career in mathematics research after completing her course and, therefore, considered the mathematical content on her course to be very important. She considered understanding concepts to be important; however, she also appeared to believe that achieving a high score in CAA tests indicated an adequate understanding.

Figure G.1: Alpha was largely satisfied that the CAA tests helped her to get the outcome she aimed for. However, she also felt that the system did not always help to identify her weaknesses.
During the first year, Alpha found her studies to be “a little bit harder” than expected, but suggested that it had a positive effect on her determination. It is clear from several parts of the interview that Alpha viewed marks from assignments as indicators of success and performance; and her pursuit of marks was a means of demonstrating the progress that was expected of her.

However, Alpha believed that CAA testing was a contradiction, somewhat, to other assessment in this regard. She noted that they did not always perform well as waypoints for measuring success: she said that CAA tests had not “been the best way to tell you how you’re doing, or even to properly test you”.

Indeed, Alpha felt that CAA did not sufficiently test students’ understanding of underlying concepts, saying, “I think if the questions for the online test were more about you having to understand certain things and certain concepts, as opposed to just putting numbers into a formula that you find, ...I think that would be more helpful.”

From her interview, it is clear that Alpha’s aims for the CAA tests was to achieve full marks. She said, “I aim for getting 100% on things like that — just because, why shouldn’t I?” Within her mathematics community, Alpha believed it was a commonly-held view that this was achievable, suggesting, “there’s a general consensus that with [CAA], it’s meant to be easier to get higher marks”, noting the “so much feedback”, ample time and opportunity to practise as reasons for this belief.

G.1.2 External influences

Alpha talked fondly about the collaboration and support that her colleagues offered to each other. Talking to colleagues in the same course, Alpha discussed her difficulty with grasping proofs and found this to be a widespread problem and there were occasions when students found aspects of the CAA tests difficult.

The library appeared to be a common meeting area for the Mathematics Department students, particular during CAA testing. Alpha had been asked for help while doing practice tests in the library. Alpha had also collaborated with a friend, another student on the same course, and believed that this was part of a supportive and helpful student community. She said, “I know that we try to help each other. At least, with my friends, we try to help each other with explaining things.”

However, while saying that collaboration during the summative testing phase was “cheating”, Alpha had been approached by students to provide assistance. She said, “What was the point, because they’ve got it now in two minutes? They’ve been watching TV while I’ve been sorting this out.”
While discussing her place in the department in relation to her studies, Alpha presented herself as a student that sought independence and wished to avoid resorting to support where possible. She was aware of the specialist mathematics support unit at the institution and that some of her colleagues had attended but she had not. Furthermore, she was reluctant to work with other students; only recently had she started working with a friend on the CAA tests. On support from lecturers, she had noted that “most of them are more than happy to spend an hour helping you” and found this surprising, adding, “I found that some of the lecturers have actually been more helpful than you’re led to believe about to do everything on your own.”

From her actions and what they said in the interview, Alpha conveyed a willingness to be somewhat independent of her peers and support services but open to engaging with her lecturers.

G.1.3 Learning Cycles

As shown in these learning cycles diagrams, Alpha concentrated on the questions she answered incorrectly in the first instance, working with a colleague more recently and compiling notes that would also assist her in the summative test. Her aim for the practice test phase was to ensure that she was able to answer every question correctly before proceeding to the summative test. In this phase, Alpha used her lecture notes and printed versions of the CAA test as well as a colleague to discuss the feedback with someone. This enabled a continuous cycle of repressing the practice tests and addressing only those questions she had answered incorrectly in the previous test. During the practice test, therefore, Alpha felt she was able to continue on the learning cycle.

However, it is important to note that Alpha placed great importance on conceptual understanding and in ensuring that she was able to carry out procedures without error. Alpha felt that CAA was weak in terms of testing the grasp of concepts — and, therefore, was not effective with regard to this aim — but was effective for refining her procedural capabilities by virtue of the comprehensive feedback offered by the CAA system. Indeed, Alpha had used a technique to obtain feedback from the system that was not wholly intended in its design.

“If there’s a question we don’t get…we go through and answer with ridiculous answers…They all come back ‘wrong’ and you get the feedback for them all.”

Alpha suggested that using this technique was a ‘last resort’ if she could not independently discover how to approach the question. Indeed, Alpha was somewhat dissatisfied that other students used this technique. She said, “I do know of colleagues who just have
(a) Alpha aimed to use the first attempt to establish which questions they cannot answer and to print the feedback for these questions.

(b) Alpha used subsequent attempts at the practice test to concentrate on questions she had answered incorrectly. She continued to use the practice tests until she could answer all the questions correctly. Alpha wrote spontaneous notes using the feedback given by the CAA system and these notes would inform Alpha’s future attempts at the questions.

(c) Alpha aimed to achieve 100% in the summative test. She would use the spontaneous notes made during the practice phase during the summative test period. Regardless whether Alpha achieved her goal to achieve 100%, there appeared to be no further path on the learning cycle.
[this] method that they’ve written down; they just put the numbers in and then that generates an answer. They haven’t really learned anything.”

The summative test, however, appears to highlight some subtle contradictions. Alpha seemed to be satisfied when she achieved full marks, but dropping marks was a cause for frustration. Alpha believed that the marks she received did not always reflect how much she understood. This occurred when she answered a question incorrectly: some questions were worth several marks and required more time and thought, but zero marks would be awarded if the final answer was incorrect.

This would suggest that Alpha placed significant important on the marks she received from the summative test, and for this phase obtaining full marks was her overriding aim. Alpha seemed to believe that the summative test was not always effective for her aim: she had diligently worked on practice tests to perfect her methods for answering the question and her subsequent errors meant she was not awarded the marks she felt they deserved.

G.1.4 Effectiveness of CAA

As the diagram suggests, the learning cycle stops promptly after the summative test. Alpha felt that the summative test served as a conclusion for CAA and that topic: in particular, she was not motivated to examine the errors she made in the summative test. However, it did spur her on in subsequent assessments; she said, “If I get a lower mark... in the tests, it just makes me go, ‘Right, in the written coursework I’m going to do so much better’.”

It is also important to consider the contradiction between Alpha’s desire to be procedurally adequate and to have a sound conceptual understanding of the topics being tested, and her dependence on the CAA system. As a consequence of becoming acquainted with the system and the techniques passed between colleagues, Alpha developed greater dependence on the feedback that the system provides as a means to answer some questions. Consequently, her development along the learning cycle lacked a progression towards autonomy. This is, perhaps, a reason why Alpha did not progress further in a learning cycle after the summative test; she did not possess the autonomy to diagnose her errors from the feedback or construct new, more challenging learning goals.

The CAA system permitted Alpha to progress along the learning cycle during the practice phase, but her dependence on the feedback and her perception of the importance of the marks proved to be limiting factors for Alpha’s potential to continue on the learning cycle beyond the summative test.
G.2 Beta

G.2.1 Identity and aims

Beta was also a student from the Mathematics Department aiming towards a 2:1 degree classification. Unlike Alpha, Beta regarded herself as requiring “ulterior motives” to work hard and quickly established during the interview that her primary motivation was marks. At the beginning of the year, Beta’s aims were more directed towards “knowledge…and experience”; however, the opportunity to undertake a placement year in Australia led to a change in her assessment aims, since she believed her marks determined whether or not she would be eligible to go.

She had also changed courses earlier in the year to drop a second subject specialism from her course since she did not enjoy that subject. This enabled her to concentrate on the mathematics and build her confidence in the mathematical module material. Beta believed that confidence was an important part of the way she studied; it proved so during CAA testing.

G.2.2 External influences

As with Alpha, Beta initially worked alone in CAA testing and her first experience of collaborating on CAA was a spontaneous one: “Me and my friend had a lecture over by or near the library…So we said, ‘Shall we just go and do the test now anyway?’ ”

Beta held similar views towards collaboration as Alpha: both were happy to engage with others while they were not being formally assessed; and both believed that summative tests should be undertaken alone. Also like Alpha, Beta had only recently adopted collaboration with colleagues towards the end of the second semester as a means for support in CAA. She explained, “I have done [that] twice. But by doing them with someone else…we’ve both gone away and done our own…There were a couple of questions that we did help each other with and check each other, but we were actually alright ourselves.”

She was less sure of the benefits of collaboration and portrayed this recent change as an experiment, after which she concluded that, perhaps, the benefits were not substantial.

“It was only the last two [CAA] tests that I did work with friends…there was [sic] three of us — we all did our own tests, then we were all stuck on one question that we weren’t sure of, so we tried each other’s and we all got different answers anyway, so we put in our own ones and we were all right. So it worked out that we were better doing it just ourselves.”
(A) Beta was primarily motivated to acquire knowledge and experience in mathematics, but she did not regard this as enough motivation.

(B) Since Beta became aware of the possibility of studying in Australia, her actions during CAA testing were more marks-driven.

**Figure G.3:** Beta’s motives changed between semesters. It had a small effect on her activity as a whole, as she tried harder to achieve 100% in her CAA tests in the second semester.
By seeking support from their colleagues, some might argue that Beta became less autonomous, rather than more so; however, seeking assistance at times of difficulty need not mean that the student was becoming more dependent. Indeed, there appear to be particular benefits for students that make use of collaboration (Black & Wiliam, 2009; Nicol & Macfarlane-Dick, 2007; Topping, 1998).

It is best to examine Beta’s CAA activity as two separate semesters, since her actions and aims were altered by the opportunity to embark upon a work placement in Australia.

**G.2.3 Learning Cycles**

In the first semester, Beta suggested she exerted less effort towards these assessment than in the second semester. This appeared to be dependent on the test difficulty as much as the change in long-term motive. When asked how many times she usually used the practice tests, she replied, “about three or four”, later adding, “there was one...we did recently that was ridiculously hard compared to ever other test that we’ve ever done and for that one, I did them probably about ten times.”

The approach in the practice tests, however, was broadly similar between the two semesters; in both cases, Beta was seeking to gain confidence in answering the questions by putting most effort into those questions she had answered incorrectly. She said “If I wasn’t confident [about the question], I would make sure the next time I did [the practice test], that I did [the question] again. And if I got it right twice, then I would be pretty sure that I’d got it right.”

Beta believed that getting 100% in the summative test was a realistic aim. She said, “I think once you actually understand the question, it’s very easy. As soon as you know how to do it, it actually becomes very easy to get 100% on the CAA.’ Beta noted that the feedback was such that working out “how to do it” was straightforward. She said that the feedback was “pretty good. When you get an answer wrong, it will tell you exactly what to do.”

However, Beta noted that her scores in the second semester were higher than previously. When asked how her CAA marks compared to other pieces of coursework she had done for the same course, she responded, “This semester, they [the CAA scores] are higher than the written courseworks, but last semester I would say they were about the same, if not lower. But I think that was just because I didn’t care as much last semester.”

This quote from Beta clearly exemplifies the change in aim and explains the transition in her actions between the semesters. It also identifies a change in the result of the
(A) Beta’s aim was to develop sufficient confidence to be able to undertake the summative test. In practice, Beta felt she had to be able to answer each question correctly to achieve such confidence. She used the feedback to develop spontaneous notes, which, in turn, would be used in the next practice test attempt.

(B) Beta embarked on the CAA summative test intent only on achieving a score, with no specific target in mind. To achieve this, she used the spontaneous notes she made during the practice tests. Once the summative test was complete, Beta gave no indication of any further use for the CAA tests.

Figure G.4: Beta’s learning cycles for the first semester.
(A) In the second semester, Beta approached the CAA practice tests much like in the first semester; however, she was more willing to engage with her peers at both the assessment and feedback stages.

(B) Beta then aimed to achieve 100% in the summative test and did so. However, she still believed there was no utility in the system once the summative test had been completed.

Figure G.5: Beta’s learning cycles for the second semester.

assessment; Beta achieved her goals in both semesters, but the goal in the second semester was more challenging.

It is not due to the assessment, however, that these goals changed. Rather, Beta found that the marks on the CAA test held greater value than she had anticipated in the first semester. This shift in aim had a minor impact on her actions during CAA, yet these changes could not be attributed to CAA itself from Beta’s comments during the interview.
G.2.4 Effectiveness of CAA

In a similar manner to Alpha, Beta appeared to develop little autonomy and remained dependent on the regime of practice tests in their preparation for the summative test. Not only did the feedback provide a method for answering the questions, but the availability of a practice test provided the motivation to practise.

Beta believed that the practice tests were worthwhile: not only for her aim of achieving 100% in the summative test, but also for broadening her experience of the techniques taught in lectures. She said, “It does actually make you do problems that you wouldn’t have done otherwise...because the last [test] that we did, that had so many questions that I had never seen before...it did widen your knowledge.” This suggested that Beta was not seeking similar experiences from elsewhere and that CAA bore sole responsibility for widening her knowledge.

Despite the abrupt halt to the learning cycle once the summative test had taken place, the practice tests appeared to be sufficiently effective for Beta’s learning goals. Like Alpha, Beta was critical of the summative test and its inability to award partial marks for incorrectly answered questions. She felt that this sometimes stopped her from achieving their goals.

G.3 Gamma

G.3.1 Identity and Aims

Gamma was a student in the Mathematics Department aiming for a 2:1 degree classification. He believed that his enrolment at university was much more than becoming a mathematics student: he was keen to experience the breadth of university life and involved himself in volunteering and charity fundraising.

He had started to find the material in his mathematics modules difficult at the time of the interview and conceded there were some modules he was enjoying less.

Nonetheless, he had maintained high expectations for himself; he said, “I usually try to get above 70% in everything — at least!” Marks were clearly important to him as he was concerned at the possibility of just missing out on a better degree classification:

“I think all those little marks count in the end and...you could miss out on a first or a 2:1 or whatever in the final exam and these little class tests.”
Gamma desired to achieve the best possible degree classification and believed that achieving 100% in the CAA summative tests was realistic. His views on collaboration and cheating were different to many of his peers.

Therefore, he considered CAA to be just as important as any other assessment and, like Alpha and Beta, believed that obtaining full marks was a realistic aim for CAA.

G.3.2 External influences

His view of the purpose of the first year had an effect on their view of collaboration and his actions in CAA testing. When asked about collaboration, he said, “I think it’s good if people work in groups, because then you can just help each other out. And if someone doesn’t understand something, then obviously everyone else can help them understand it, especially in the first year when it seems to be about getting everybody to the same sort of level, because it doesn’t actually count with the marks.”

In that quote, Gamma revealed three of his views that had an important effect on his actions in CAA:

- he believed that the purpose of collaboration is to ensure everyone in the group attains a certain level of understanding;
- he believed that the purpose of the first year at university is to ensure that everyone shares a common level of understanding to progress into subsequent years; and
• he believed that the marks gained in the first year do not contribute to a student’s degree classification and so plagiarism is not as big an issue as it would be in subsequent years.

These beliefs extended to the summative test. Gamma said, “On the [summative] tests, I think it’s not cheating to help each other out, really. It’s just offering help, really — I don’t know — trying to get everyone to understand it, because I think the whole point is to get it so that everyone understands the content.”

Gamma remained cautious when discussing how he engaged with colleagues during CAA. The only indication that he, personally, had collaborated was when he was asked how an invigilated test would change his actions. He replied, “I wouldn’t be able to ask my friends to check my answers that I wasn’t sure about.”

G.3.3 Learning Cycles

Gamma’s approach to the CAA practice tests was similar to Alpha’s approach. Both were keen to ensure they were able to answer all the practice test questions before proceeding to the summative test. Gamma said, “I might end up doing the test seven or eight times, but I wouldn’t have done all the questions every time. I would just do the ones I kept getting wrong, just to make sure I could get it right.”

Like some of the other students in the study, Gamma would make a serious attempt at the practice test before printing off the feedback and using the resulting document as a basis upon which to model and improve his answers.

As noted previously, it was not possible to establish how Gamma engaged with collaboration with certainty. He gave indications that not only was he in favour of working with others, but also that he may have worked with others in the summative test.

Once Gamma was confident that he could answer all the practice test questions, he was sufficiently confident to take the summative test. In preparing for the summative test, Gamma refined his notes and collated the CAA system feedback that he had earlier printed. He would use the library or a computer laboratory to undertake the summative test, since he was concerned of possible malfunctions with his own computer; Alpha noted that the library was also a common meeting place for students wishing to do the CAA summative test.

Gamma, like Alpha and Beta, gave no indication of inclining to correct errors after the summative test. Although Gamma was wary of becoming too dependent on the question
(A) Gamma’s approach was similar to Alpha’s: the aim of the first attempt at the CAA practice test was to establish which questions he was unable to do and to use the feedback provided by the system to model his corrections.

(B) Gamma used the practice tests to perfect the questions he kept answering incorrectly. The feedback would be used to inform his future methods until he was able to answer all the questions correctly.

(c) Gamma implicitly gave the suggestion that he worked with peers during the summative test. Where Gamma differed from Alpha was after the summative test was completed: regardless whether Gamma achieved his aim to gain full marks, he planned to use the assessment for his next goal.
formats and contexts presented in these tests, he was interested in using the practice tests as preparation for their end-of-module exam:

“I don’t know if the sort of questions that come up in the [CAA practice] test would be similar to the exam-style questions... I probably would use the computer things [CAA practice tests] as well.”

G.3.4 Effectiveness of CAA

In judging the effectiveness of the CAA system for Gamma, it would appear his learning cycles are similar to those of Alpha and Beta. Gamma was largely motivated by the pursuit of marks since, in his view, marks were necessary to ensure that he would obtain the degree classification he was aiming for.

Gamma felt that the practice tests enabled him to achieve his initial learning goal: to be able to answer all the questions correctly before going into the summative test. For the summative test, it is unclear from Gamma’s comments whether he felt he had achieved his learning goal — to achieve 100% — on a regular basis. Nonetheless, Gamma continued on the learning cycle by revising his learning goals with respect to his exam revision after the summative test. Since the interview was held before the exam, it was not possible to determine whether Gamma used CAA to revise and whether it was effective for that aim.

G.4 Delta

G.4.1 Identity and Aims

Delta was a joint honours student in the Mathematics Department and the computer science department. She was aiming to achieve a first class degree. Though she was progressing well with the computer science modules, Delta found the mathematics modules tougher than anticipated. Consequently, Delta sacrificed aspects of her social life to keep up. She said, “I don’t really go out much. I have a lot of work to do. That’s pretty much my life, just trying to get my grades up.”

When asked about aims for the course, Delta was explicit in her pursuit of marks. She aimed for higher scores in CAA than in other assignments:
Delta viewed marks as being the reward for effort and perseverance and, as such, was prepared to work hard to achieve her learning goals.

G.4.2 External influences

Unlike her mathematics peers in this sample, Delta was very much against collaboration. She said, “I don’t like collaboration as it is, anyway, in any coursework because I think that people who are good at it [the subject matter] will take the majority of the work, and then people who aren’t really sure just sort of survive off their work.” She believed that plagiarism was a problem. She said, “They mooch off other people to do their coursework for them, and they get the marks when they don’t deserve them.”

As such, she believed that the CAA system was unfair, since the system could not distinguish between those that work independently and those that do not. Despite her views, Delta did engage in some paired work as she felt this arrangement was beneficial to both parties, saying, “I think it’s ok if you work in a pair, because sometimes it is good just to bounce ideas off someone and to check that what you’re doing is right.”
(A) Before embarking on the summative test, Delta aimed to ensure that she was able to achieve 90% in the practice tests. Doing so meant that she was able to achieve 90% in the summative test. Like the other mathematics students, she paid attention only to those questions that she answered incorrectly. Delta noted that she engaged with peers during the practice tests and that she noted the methods used in the CAA feedback.

(b) Delta aimed to achieve 90% in the summative test. Like Alpha, Delta did not perceive any use for CAA after the summative test.

**Figure G.9:** Delta’s learning cycles.

### G.4.3 Learning Cycles

When undergoing the CAA practice tests, Delta had an ideal score of 90% in mind. In order to achieve this, she would take an initial test to establish which questions she was not able to answer correctly and receive the feedback. Like others in her cohort, Delta focussed her efforts on the questions she answered incorrectly. However, unlike her peers, Delta attempted the entire test each time.

To address those questions they answered incorrectly, Delta would immediately adopt the methods presented in the feedback. She said, “I’ll use the example they’ve used and I’ll just work through the question [in] exactly the same [way] — so, how they want you
to do it.” That is, Delta felt she was learning the processes that she was expected to know by mimicking the feedback.

The practice test would last until Delta achieved the score she wanted. She said, “When I’m consistently getting the number [score] I want, then I’ll take the real [summative] one.” This would sometimes take numerous attempts: “if it takes…seven practice tests to get 50%, then I’ll be like, ‘Oh, I need to do some more’.”

Despite these efforts, Delta said she did not take the tests particularly seriously, in comparison to other coursework types. She said, “No, I don’t work harder towards CAA. if anything, I don’t treat it like a proper coursework…I don’t think they’re as serious.”

During the summative test, Delta worked in the library and had her lecture notes and the notes she had composed during the practice test. Working alone, she had a mixed experience in terms of results. Delta believed that “it’s easier to achieve higher in them”, but they also said, “I’ve got 100% on a couple and then I failed a couple.”

Delta believed that this was partly due to the feedback. There were occasions when she was unable to make a start on a question because the feedback did not sufficiently help her. She said, “They’re [the CAA questions] just very repetitive, and I obviously get frustrated easily. So if I still can’t do it — if I just really, really can’t do it — I just get to the point where I want to give up, because obviously no-one’s telling me where I’m going wrong.”

This indicated a reliance on the CAA system for providing the feedback that she needed; Delta was aware of other support channels, including a support centre for mathematics difficulties, but did not use them. When Delta reached that point, she did not express a willingness to compromise her initial learning goal; it would appear that Delta continued to the summative test regardless and evidently failed the test on occasion.

Perhaps because of these failures — although other students voiced similar frustrations — Delta believed that the marks she had been awarded did not adequately reflect her performance in the test. She said, “You don’t get marks even when your working out was right and you just messed up a number somewhere.” For Delta, frequent minor mistakes were costly: “I often have a tendency to leave square roots out and make silly mistakes like that and it doesn’t pick upon that, obviously.”

G.4.4 Effectiveness of CAA

When examining the effectiveness of CAA for Delta, the rigidity of her goals and her subsequent problems in reaching those goals meant that it was somewhat inevitable
that sometimes CAA would not be effective. Although she displayed concerted effort in order to meet those goals, Delta’s dependence on the feedback to “learn” how to do the questions meant that when this procedure did not work, Delta did not opt for further support.

Partly because of her failures and partly due to CAA testing procedures rather than concepts, Delta said she felt “stupid, because I know I can do maths”, yet she said, “You have examples of how to do it [in the feedback], so you just work through a step-by-step guide.”

Thus whenever Delta came upon a problem she was unable to tackle, CAA was no longer effective — indeed, Delta noted that without further intervention from the system, “You can’t learn from it” — and required support from elsewhere. She made the comparison with written coursework, where “your tutor can actually point you to where you’ve gone wrong.”

Delta’s stance on collaboration, therefore, became significant. She was, for the most part, unwilling to accept help from her peers. She discovered later in the semester that working with another student had benefitted her. When this happened, Delta was able to provide and obtain feedback from another source outside the system.

Beyond the goals for the summative test, Delta’s learning abruptly halts. Since the system did not provide detailed feedback at this stage, she felt unable to learning anything further from the assessment — further indicating her dependence on the system feedback. Furthermore, Delta believed that summative assessment generally represented a closure at the end of a topic.

Although Delta was keen to be independent and to develop autonomy, it was this desire that prevented her from seeking help at times of difficulty. Ultimately, this had a detrimental effect on her learning and limited the scope of CAA to be effective for her.

G.5 Epsilon

G.5.1 Identity and Aims

Epsilon was an Engineering Department A student aiming to achieve a 2:1 degree classification. He believed himself to be hard-working — particularly in comparison to other students — but was keen to enjoy the breadth of student life, much like Gamma.

With regard to mathematics, Epsilon said that mathematics was one of his strongest subjects and thus aimed to “get good marks in maths”. His view of mathematics was
Epsilon was motivated towards marks and understanding the material in his motive to become a pilot. No obvious contradictions arose in his interviews. regarded through his identity as an engineering student; he said, “If you do muck up... in real life... you kind of only get one chance.”

Epsilon was the only student in the sample to have explicit learning goals that transcended assessment and were related to his motive to become a pilot. He said, “I hope to gain more of an understanding of how aircraft are built and how they are operated.”

This motivated his quest for understanding — as well as obtaining marks — and viewed success as a combination of acquiring the two. He said:

“I quite enjoy maths as a subject as well and understanding it, because I think it’s quite fundamental to how everything works. So as well as getting the grades, I think it’s really useful [to understand] the subjects, because then you can apply it to real-life situations and other engineering examples as well. So that motivates me as well.”

Since Epsilon needed an upper second class degree to become a pilot, he was aware that this had an impact on the grades and marks he aimed for in his modules and assignments. He referred to “the big 70%” and both the desire for understanding and the need to achieve certain scores encouraged a pragmatic approach. He said, “I obviously want to do well in university and get a good degree out of it. I want to try to get the best I can out of every single module, so I do not do well in other modules, subjects like maths will bring my average up.”
G.5.2 External influences

Epsilon believed that collaboration was much less the norm in Engineering Department A than in other departments. He was in favour of collaboration and worked with two colleagues during CAA testing.

He was clear in his belief that collaboration was not appropriate for summative tests, and in CAA summative tests in particular. However, in his view, collaboration in the practice tests offered the opportunity for learning, leaving the summative test to test that learning.

G.5.3 Learning Cycles

Epsilon began each practice test phase by attempting the tests alone, using his lecture notes and the module workbooks. He used the feedback to establish where he had gone wrong. He believed that the CAA practice tests worked well with the workbooks and the lecture examples to gain a good understanding of the types of questions he might be expected to answer, both in the summative test and in the exam.

After one or two attempts on his own, Epsilon collaborated with two other students on the same course:

“There’s a couple of [students] in my halls who do the same course. I normally try once or twice by myself to see where I’ve gone wrong, and then they’d normally do it as well. Then, we compare with each other afterwards, so we usually see where each other had gone wrong and help each other out. I think that’s really useful and helpful.”

Epsilon used the combination of lecture notes, workbooks, the practice tests and his friends in order to prepare for the summative test. In one semester, the paper test was marked as though it had been a computer test conducted ‘offline’, that is, students were awarded scores only for correct answers with no further feedback. In the other semester, marks were awarded for the application of the correct method, even if mistakes were made, and Epsilon found it easier to identify where he could improve.

His aim for summative assessment was in terms of scores; he had used the term “big 70%” to describe this aim. Epsilon appeared to achieve his aim each time, but what happened after the summative test depended on the style in which the paper test marking was conducted. Epsilon was more inclined to check for errors if more feedback was given.
(A) Epsilon used the practice tests a few times initially to establish how much he already understood. He took notes from the feedback but did not use the tests to improve performance at this stage.

(b) After performing the practice tests a few times initially, Epsilon worked with peers to address the gaps in each other’s understanding. Once everyone in the group was able to answer all the questions correctly, they considered themselves ready for the summative test. He worked with many resources and composed a document with parts of the CAA feedback he thought would be useful.

(c) In the summative test, Epsilon aimed to achieve 70%. Since the test was invigilated and paper-based, there were no additional resources that Epsilon used. He achieved his goals, but did not envisage any use for CAA beyond the summative test.

**Figure G.11:** Epsilon’s learning cycles.
Epsilon stated that making small errors was frustrating — particularly so in those tests in which making such mistakes was interpreted as answering the question incorrectly. However, he remained pragmatic, saying, “I’m not too keen on the fact they don’t give you the working marks for it, but then again, in the real world, you’re not always going to get the opportunity [to make mistakes].”

G.5.4 Effectiveness of CAA

For Epsilon, the CAA practice tests seemed particularly effective. By working independently to start with and addressing the problems he was able to — before consulting with colleagues — he displayed autonomy and responsibility. His aim for the practice tests was to gain an awareness of his own understanding before attempting to consolidate and build upon this understanding with the assistance of colleagues.

In the summative tests, Epsilon found the paper test with detailed feedback led to opportunities for further learning and, thus, it was more effective. However, Epsilon also found the paper test without additional feedback to be reasonably effective and considered having a paper test to be easier to manage than the computer-based test:

“I find it a lot easier with the questions written down on a piece of paper in from of me, because I can highlight bits, scribble over it, pick out important parts of it. I find it a lot easier to work not staring at a computer screen.”

Like some other students, Epsilon also anticipated using the practice tests as preparation for the module exam.

G.6 Zeta

G.6.1 Identity and Aims

Zeta was a student in Engineering Department A and was pursuing a 2:1 degree classification. Zeta and Epsilon appeared to be friends, since they sat next to each other when completing the questionnaires and were interviewed consecutively. At the time of the interview, Zeta was undecided in what he hoped to do after his course; his aim was to obtain “a degree — I see it as a step to getting to different places.”

In terms of assessment — and CAA in particular — he appeared to view interim assessment as being opportunities to identify areas in which he was weaker prior to the exam.
For these assessment, he did not have a specific aim in mind, though he said, “I like to win, or I like to aim high, so I like to get the green ticks in the little boxes [in the CAA practice tests].”

G.6.2 External influences

Zeta’s stance towards collaboration was in stark contrast to Epsilon’s, however. Zeta viewed collaboration in CAA as unnecessary since the standard required of students to embark on this course should be sufficient enough to make the first year mathematics modules straightforward. Furthermore, he felt that the feedback offered by the CAA system was detailed enough so that there would be little gained from discussing the work with colleagues.

Despite his views on collaboration, Zeta willingly offered support to those that needed it on occasion. He cited the benefits he received in the understanding of the material when “teaching” others. However, he usually worked alone — “I’d say 90 to 95% of the time.”

G.6.3 Learning Cycles

Zeta began each testing phase by writing out the questions on paper. He said, “It allows me to obviously first of all see what I need to do, and then it gives me time to look through my book to find notes...then I work through it and then put an answer in.”

During the practice test phase, he had his lecture notes, his module workbook, and his
(A) Zeta worked in a similar fashion to Alpha and Gamma. To begin the practice phase, he would make one attempt at the questions to establish which questions he could not answer correctly. His aims for this initial attempt also included obtaining the CAA feedback for modelling corrections.

(b) For the rest of the practice tests, Zeta used the feedback to improve the incorrect answers until he was able to answer all the questions correctly. As well as the feedback, Zeta used his lecture notes, workbooks and the formula booklet.

(c) Zeta’s aim for the summative test was to achieve high marks and always achieved his aim. However, there was no indication that Zeta had any intention to use CAA after the summative test.

Figure G.13: Zeta’s learning cycles.
handwritten notes. Although Zeta did not mention so in the interview, he said in the questionnaire that he would usually attempt the test five times or more.

Zeta found — like other students — that it was possible to obtain the feedback from the questions without making a serious attempt at answering the questions; however, he was reluctant to do this too frequently: “the only time that I do it is if I have no clue — if I see a question and I have no idea what’s going on, that’s the only time [I do that].”

Like many of the interviewed students, Zeta gave suggestions of dependence on the feedback. Generally, he found the feedback useful and it helped him to improve. He appreciated the variety of feedback, which sometimes included full-worked solutions and sometimes referred students to a particular reference point in their workbooks or lecture notes. In the latter cases, more effort was required from the student to correct their mistakes; however, Zeta found this to be frustrating. He said, “There are some where it’s just the theory, which is ok, but I guess it requires a bit more effort and that’s just stupid.”

In the summative test, Zeta aimed to get “high marks” — or between eighty and ninety per cent, as suggested in his questionnaire response — and tended to achieve that score. Indeed, he was surprised that his marks did not change significantly between the two semesters, given that in one semester methods marks were not awarded.

Zeta, like others, found the lack of method marks to be unusual and was very much in favour of gaining credit for applying a method correctly: “I’ve got quite a strong opinion on these method marks. I’m sort of one of those students that I’ll do all the right working out and I’ll get everything right and just mix up a number at the end, which [is] giving me zero.”

Like Epsilon, Zeta appreciated the learning experience arising from receiving zero marks, saying, “It is effective on the alternative argument that, in real life, you have to get the right answer.” However, he found receiving zero to be a disincentive for correcting mistakes, suggesting that a zero mark does not indicate the size of improvement required.

### G.6.4 Effectiveness of CAA

Nonetheless, the combination of the CAA practice tests and the paper summative tests appeared to be effective for Zeta. For the most part, CAA practice tests enabled him to identify his weak areas, which had a purpose for both the summative test and the exam. The summative tests offered some learning experiences and Zeta tended to get the score he was aiming for.
Zeta remained dependent on the feedback offered by the CAA system. It did not seem to have a detrimental effect on his ability to achieve his learning goals, however, and there were evidently instances where Zeta followed the advice of the feedback to seek assistance in his lecture notes or workbooks. He displayed some autonomy in assisting others and, Zeta insisted, these occurrences helped him develop a greater understanding of these topics.

It is also unclear what Zeta did after the summative tests in each case, and thus whether Zeta continued on the learning cycle beyond the summative test is not revealed. He suggested he did not seek to correct mistakes in the summative test — particularly for questions for which they received zero marks — but he was considering using the practice tests for exam revision.

### G.7 Eta

#### G.7.1 Identity and Aims

Eta was an Engineering Department A student that aimed for a first class degree. Before university, he studied Further Mathematics at A Level and this gave him confidence and experience for the first year material, saying that the modules were “relatively easy”.

He characterised himself as a student that was able to learn new material quickly; however, he was prone to distractions and procrastination. He aimed to achieve 100% in every test in order to acquire the full 5% of the overall module score given to CAA testing.
Like Zeta, Eta did not have clear intentions for what he wanted to do after university, though he indicated an interest in working in engineering in the questionnaire.

G.7.2 External influences

Eta believed that collaboration was quite widespread among the Engineering Department A cohort — somewhat at odds with Epsilon’s beliefs. When Eta fell ill towards the beginning of the second semester, he depended on the support of other students. He presented a positive impression of the group of students that assisted them:

“My friends did come over to help me figure out the inverse Laplace and all that. In my block, there’s about four of us who do the Aero/Auto degree and we all have to do the maths tests. At one point, there was [sic] all of us in the room on the whiteboards trying to figure out one of the questions. The group work does help. It’s quite useful just to see how other people do it, see if there’s a quicker way to do it, comparing it to the formula book and all of that. But yeah, that’s good for when you want to learn how to do the method quicker. Sometimes, on your own, you don’t really get it. But when you’re in a group, it just all comes together.”

Not only did Eta suggest that such collaborations were commonplace, but group work was particularly well-organised. Many groups took to seminar rooms to discuss methods and questions on the whiteboards for coursework and lab reports, as well as CAA.

Though Eta found group work particularly helpful for catching up, he did not always find it useful. He said, “At times, if people work at a fast or slower pace than you, I find it easier to go away by myself and do it in my own time.” Consequently, he tended to work with others at the start of the practice test phase before working independently to ensure that he had perfected the techniques.

G.7.3 Learning Cycles

Eta had developed a regime for the CAA practice tests. Initially, he would read his lecture notes “because it’s just copied off the board. I don’t get time to understand”. He would then work with colleagues and discuss any problems they had encountered. After trying the tests with others, Eta would “try to do the methods by myself and then perfect them.”
(A) Eta’s practice regime was somewhat the reverse of Epsilon’s. Eta worked with peers first to discuss the assessment and the feedback to identify the questions he cannot do. Peer support provided additional feedback to that provided by CAA. After a period of group work, Eta changed his aim to ensure that he was able to answer all the questions independent of his peers.

(b) Once working independently, Eta’s learning cycle is similar to many of the other students: Eta continued to use practice tests until he was confident that he could answer all the questions correctly, and used the CAA feedback to create spontaneous notes. Unlike the other students that adopted similar cycles, Eta answered every question every time he took the practice test.

(c) Eta aimed to achieve 100% in the paper test and suggested that he often achieved this aim. Unlike the other students in this study, Eta desired to develop further understanding after the summative test, but the learning cycle ended here. There were two reasons for this: Eta felt pressured for time and prioritised the work that would give them marks; and Eta felt that the CAA system was less well-equipped to test students’ conceptual knowledge.

Figure G.15: Eta’s learning cycles.
When working independently, Eta would attempt every question and concentrate more time on those he answered incorrectly on the previous attempt. Unlike others in this sample, Eta stated explicitly that he answered every question on each attempt at the practice test. These iterations continued until he was confident that his methods for each question were correct.

For the summative test, Eta felt prepared having done the practice tests and workbook questions. He noticed that these two resources complemented each other and, together, prepared him sufficiently for the summative test. He said, “They definitely do help. I couldn’t do well without them really — the [workbooks] or the practice tests.”

Although some other students mentioned that the CAA feedback was not personal, Eta believed that, because the feedback was not personalised to his responses, it should not be considered true feedback. Comparing CAA to other types of coursework, Eta said, “Detailed feedback makes coursework very useful…You know when you do the coursework that [the feedback tells you], ‘I need to practise this’…So, if it [CAA] had feedback, that would make it much better” (emphasis added).

Beyond the summative test, Eta believed there was potential in the system to enable him to further develop his understanding. However, Eta felt that had the system possessed this capability, he would have been tempted to overlook more challenging questions due to time constraints. He suggested that time pressures were a reason for not dwelling on CAA once the summative test had been marked.

G.7.4 Effectiveness of CAA

For Eta, the system had been relatively effective. During the course of the practice test phase, he developed greater autonomy by ensuring he could complete every question without his peers, albeit with assistance from the feedback, workbooks and lecture notes. Unlike some of the other students in this sample, Eta displayed some flexibility and ability to address problems from various contexts by not relying on the CAA feedback alone.

Furthermore, Eta appeared to achieve his learning goals for both practice and summative phases. However, Eta progressed further than most other students in the sample insofar much that this testing regime inspired the creation of a new, more challenging learning goal. Unfortunately, he found that CAA was not effective at testing these new goals and he would not have sufficient time to develop the understanding he desired to acquire anyway.
G.8 Theta

G.8.1 Identity and Aims

Theta was an Engineering Department E student who declared that she had not considered a particular degree classification to aim for. Prior to arriving at the institution, Theta had been working in materials and in engineering roles. She had used little mathematics in that time and had not studied mathematics in her post-16 education. Therefore, there had been some years since Theta used mathematics regularly and had not studied it to the same level as many of her peers.

This led to Theta feeling a lack of confidence; she was aware that most of her colleagues had greater and more recent experience of mathematics. Nonetheless, she confirmed that she was enjoying the course, saying, “The struggle’s been worth it.”

Her aim was to improve her employability and, as such, she aimed to get a degree. For the mathematics components of the course, Theta aimed to pass and then to do as well as possible beyond that. Because twenty per cent of the module mark comprised CAA scores, Theta saw the tests as an opportunity to gain marks prior to the exam: she aimed to have “bagged a good 15%”.

G.8.2 External influences

Theta believed that collaboration was uncommon among the cohort of Engineering Department E students in the mathematics modules and was aware that collaboration...
within the cohort was more common in the Engineering Department E modules.

When asked, Theta was open-minded with respect to collaboration. Initially, she had reservations about it and worked alone; however, an obligatory collaborative assignment in another module challenged her preconceptions and she found the exercise enjoyable. However, she placed great importance on understanding the material and felt that too much collaborative work would endanger the breadth of her experience.

Theta believed she was independent as a learner, and appreciated that this did not mean working alone. She was prepared to ask for help when it was required. There were occasions in CAA when she requested support, but ensured that she was able to solve the problem independently afterwards.

G.8.3 Learning Cycles

Theta approached the practice tests in a way similar to some other students in the sample: she would take a practice test in the first instance to identify problem questions, then concentrate on those questions she had answered incorrectly. Theta also gave much clearer detail over her regime than other students, saying, “I think it goes up the week before, so I’ll practise three times a day, every day, up until the day [of the summative test]. It’s usually sat on a Thursday, so I’ll practise, like, five times through the day so it’s fresh in my head, ready for the test.” Theta used the practice tests to ensure she was confident enough with the material to pass the test.

In the summative test, Theta’s aim was to do as well as possible but, at the very least, her priority was to pass the test. As in the case of the Engineering Department A cohort, Theta experienced different CAA practices between the two semesters: in one semester, students completed invigilated, open-book paper tests; in the other, the tests were also invigilated but conducted online on the CAA system, and no additional materials were permitted. Theta found the transition between the two to be a challenge.

In the first semester, Theta was able to achieve her goals. The summative test questions were similar to the practice test questions so she felt adequately prepared and confident. In the second semester, the questions demanded the same techniques but of unfamiliar contexts; and Theta found it difficult to adjust:

“The open-book exam made me think, ‘Well, if I can get the answers right in practice, there’s no reason as to why something similar like this wouldn’t come up.’ So obviously, I didn’t write down the answer, but I would write down the method that I had used. And I got to the real [open-book] one
(A) Theta aimed to develop her confidence and understanding of the topics, since she felt that was where she was weak compared to their peers. The way in which she tackled the practice tests was very similar to some of the other students in this study. Theta would continue to use the practice tests until she was sufficiently confident for the summative test, concentrating on answers she had answered incorrectly in the previous cycle. Theta also used the mathematics support centre for assistance when she encountered difficulty.

(B) Theta’s aim for the summative test was to pass. Since the test was invigilated, she did not use any external resources. Nonetheless, she achieved her aim in every one of the practice tests in the first semester. After the summative test, Theta intended to use the practice tests to revise for the exam.

Figure G.17: Theta’s learning cycles for the first semester.

and the methods didn’t help, because there were other methods that were needed. ...I was so confused that in the end I just thought, ‘Ok, I’m just going to have to accept the fact that I could fail this one and be ready for the next one’.”

G.8.4 Effectiveness of CAA

The second semester demonstrated the problem of dependence on the CAA system: Theta remained dependent on the contexts presented in CAA and, therefore, struggled
(A) Although Theta was aware of the difference in the CAA practice adopted by the two lecturers, Theta approached the practice tests in the same way in the second semester as she did in the first semester.

(b) The paper test was invigilated and open-book, and the context of the questions were less familiar to those that were accustomed to the questions on the CAA practice tests. Theta felt that the unfamiliarity of the context caused her to fail some of the paper tests. When Theta passed the tests, she expressed the same desire to use the tests to revise for the exam, since she believed they would be useful; however, when she failed the test, Theta’s learning cycle stopped.

**Figure G.18:** Theta’s learning cycles for the second semester.

when the context changed. It highlights a problem that might not be observed elsewhere, since in the other cases in this sample, the contexts used in questions remain largely consistent between the practice tests and the summative tests.

Unfortunately for Theta, she failed some of these tests, so she did not achieve her goal of passing the tests. Furthermore, at this point Theta decided to abandon that topic and concentrate her efforts on the next test. Consequently, this demonstrated that the testing regime was not so effective for Theta, as her learning cycle stopped at this point.

As for autonomy, Theta display some self-regulation skills in how she conducted her
learning. She avoided dependence on peers for support, yet realised when it was appropriate to seek help. However, this autonomy led Theta to believe that she was adequately prepared for the test. Thus it would appear that CAA was not effective at assisting Theta to develop sufficient autonomy from the system itself.

However, like some of the students, Theta expressed a desire to maintain access to the practice tests as preparation for the module exam.

G.9 Iota

G.9.1 Identity and Aims

Iota was an Engineering Department G student that was aiming for a 2:1. Like Theta, Iota considered his mathematical background to be inferior to his peers, having not studied mathematics beyond the age of sixteen. Although he “scraped the requirement” for the course in terms of mathematics study, Iota said of the mathematics component of the course, “It’s not been too bad, to be fair. We do...some complex stuff, but they do start at the beginning of every topic with the basics.”

Iota was clear with his aim: “I’m at university to get a degree, to get the best degree I can possibly get. I’m in it for 100%”. He viewed CAA as being part of a “pre-exam block”, which comprised summative assessments and coursework that are completed prior to the exam. In Iota’s mathematics module, the pre-exam block was worth 40% of the overall module score, of which half was CAA. Therefore, Iota viewed CAA as an opportunity to account for 20% of the module score before the exam.

He believed that this notion was widespread across the Engineering Department G cohort. The fact that the pre-exam block was 40% was significant: the pass mark for the module was 40%, so students could be reasonably assured they would pass the module if they performed well in these assignments. Iota said:

“I think I’m on about 17 or 18 [per cent] now, plus the potential 20% from the coursework, which should be about the same — you don’t need more more to bump it over to a pass grade...to have up to 40% before the exam is pretty significant in my eyes, I would say. Anything you can get before just helps to take pressure off that exam situation.”
However, while marks were Iota’s primary aim, he was aware of the need to develop understanding since the material in the first year mathematics course would be a foundation for new material in subsequent years. He felt that he was catching up with his peers in these terms, adding, “And that’s obviously through the [CAA].”

G.9.2 External influences

Though Iota enjoyed experiences of collaboration, he believed it was because he was with “like-minded people” with similar aims — “to get the best mark they can” — and was wary of being placed with others with different aims.

Despite the positive experiences, Iota did not collaborate while using CAA and believed that collaborating on CAA was rare in the cohort. He explained, “I think...for most people on my course, the CAA tests are a bit of a walk in the park, because I think the majority have done maths to a high level, so it’s mostly revision to them. So they just arrive, sit down, and ten minutes later, they’re gone.”

G.9.3 Learning Cycles

Iota found the regime of practice similar questions multiple times to be helpful and felt that it was particularly beneficial given his mathematical experience. He also felt that the opportunity to learn by “doing” was compatible with his learning style. He said, “I tend to absorb the theory really by learning through applying it, and that’s beauty of this CAA system where I can just go through and do the questions over and over again.”

When doing the practice tests, Iota would have his lecture notes, formula book, a calculator and anything else he deemed potentially useful and conducted the tests in his room. He did the test several times in order to ensure he had grasped the method. He said, “Even if I’ve done the question ten times and I’m sure how to do it, I’ll make sure that I go through [the whole test] writing the whole method, just so it sinks in.”

Like Theta, Iota made use of the availability of the practice tests over an entire week and was very happy with the feedback it offered: “you get your overall mark and then which questions were right and wrong...the wrong answers, it shows your answer and then the working and method to get to the right answer, and any general results you might need to get there.” Furthermore, the immediacy of the feedback meant that he did not have to wait to observe improvements: “You’re not waiting two weeks and you’ve forgotten what the questions were and what you had written.”
When it came to the summative test, Iota had a set of written notes containing things he considered important or useful, as well as the formula booklet. He did not have a specific aim and, despite insisting that marks were important at the start of the interview, Iota seemed relatively content to have acquired sufficient understanding to be at a par with his peers. Consequently, Iota felt that CAA was more testing for him than for his peers.

G.9.4 Effectiveness of CAA

Like others, Iota felt the lack of a human assessor results in the loss of an opportunity for receiving feedback — and marks: he said, “If you were to take a human-written exam, they can obviously see that you’ve written all your working and you’re going to get two out of four [for a sign error] or whatever, it may be just for the working...but if you put it into the CAA, you get nought out.”

Judging Iota’s reflections on CAA, it would appear that CAA was effective for his aims to reach a similar level of understanding as his peers. As for marks, he was pleased with having gained 17 to 18% in the pre-exam block from CAA.

For the summative test, Iota’s aim was not particularly challenging, and perhaps it indicates that he did not anticipate the possibility of learning taking place in a summative test. It also resulted in a halt to the learning cycle once the summative test was completed. With respect to autonomy, Iota remained somewhat dependent on the system to provide opportunities for practice; however, he also found that this system was perfectly sufficient for his aims.
Appendix H

Lecturer interview case studies

H.1 Kappa

H.1.1 Subject

Kappa was an experienced lecturer whose research was in the field of mathematics pedagogy. She reflected upon her teaching practice in light of her own research and her reading of others’ contribution to mathematics pedagogy literature.

She was regarded highly for this expertise both within the department and more widely in the mathematics education community. As Roth (2007b) argued, identity is a construction from what an individual has done and what they think. In Kappa’s case, her

\begin{figure}
\centering
\includegraphics[width=\textwidth]{kappa_activity_diagram}
\caption{Kappa’s activity for assessment showing contradictions sited within the subject and between the tools and object of activity}
\end{figure}
research, her practice and the beliefs that she had generated through her work comprised her identity and moulded the way she acted when teaching.

Using the notion of figured worlds (Holland:2001wn), Kappa had developed identities as a teacher, researcher of mathematics pedagogy and expert in mathematics education. These figured worlds were by no means disparate: they were related and were mutually influential. For the activity of assessing her students, Kappa spoke from each of these identities, demonstrating the influences that they have on each other: in particular, she noted the introduction of the group projects as a result of a research project she had been working on.

H.1.2 Object

Kappa’s object for the activity was driven by her research and reading. She emphasised the need to allow students to explore the concepts behind the mathematics they were using, which needed to occur alongside giving students the confidence to perform the procedures they are required to learn. Therefore, she realised the importance of giving students the opportunity to practise and gain feedback on their work.

By offering regular interim assessments, Kappa could monitor students and allow students to monitor themselves in their learning. However, such regular testing would be a burden on her time in preparing and writing meaningful feedback using traditional assessments techniques. While she wanted to offer students this facility, she acknowledged that she was already pressured for time.

H.1.3 Tools

She used CAA with her students in Engineering Department E. In the past she had been reluctant to use some online assessment resources as she found them to lack rigour and had been rudimentary. She had been generally keen to introduce technology in her teaching where it would help students to grasp topics and develop their conceptual awareness.

She used the CAA system with these students since they offered the opportunity for students to practise: “I think that they fulfil a very useful purpose in enabling students to practise and become familiar and confident with the way of doing things.”

This was not the only aim that Kappa had for the interim assessments. As well as offering the opportunity to practise, Kappa desired to test students’ conceptual understanding: “I put an important emphasis on conceptual work, but that doesn’t say there’s no need
Appendix

Kappa felt that the CAA system was not sufficient for testing conceptual understanding in its current state, but fulfilled the purpose of ensuring that students were tested on their procedural work.

For these reasons, Kappa found it insufficient to depend on one tool to carry out the activity and adopted other assessment techniques to reach the object of the activity. While CAA was a useful tool for testing many students quickly, other assessment tools could be used to test conceptual understanding. She said, “If they [the CAA tests] were the only form of assessment, then I would feel that maths learning was impoverished, because there is a lot of encouragement for instrumental learning.”

Kappa was keen not to reduce the assessment regime to one kind of assessment tool; she was content to have the variety of CAA alongside projects to assess different types of knowledge of understanding. For that reason, she was not keen on the idea of introducing questions that assess conceptual understanding to CAA: “It’s not so much that I want to put more challenging questions on the CAA, but by themselves, they [the CAA tests] don’t offer so much of a challenge as I want. But given that the CAA is part of a wider assessment programme, then the CAA alone is not offering the degree of challenge I want overall.”

H.1.4 Community

Kappa was a member of many communities that had an impact on her activity of assessing her students. She was a member of the mathematics lecturers and department; she was a member of local, regional and international communities of mathematics education researchers; and she keenly kept abreast of innovations in mathematics learning, pedagogy and assessment literature.

These communities affected her in many ways: not least in the formation of her leading identity for this activity. She was aware of the various tools at her disposal that would help with her activity: from the CAA made available by the department, to the techniques discussed in the other communities.

Division of labour

One of the many reasons that lecturers are tempted to use the CAA system is the opportunity to save time by having assessments marked automatically by a computer. Further time is saved in the preparation of such assessments since a staff member is dedicated to setting up tests to the lecturer’s liking. In practice, this meant that lecturers
could request which questions appear in the tests, how long the practice tests are made available, and precisely when the summative test is made available.

Kappa made use of this division of labour: “What’s particularly good about them… is that more or less somebody else does all the work. So, I get a form of assessment that is useful within the boundaries that I’ve stated, and it happens. Whereas for the projects [which assess conceptual understanding], I’ve just spent almost the entire weekend marking a set of projects… They are hugely labour-intensive, whereas CAA tests are not.”

The workbooks, which are given to students as a core document for each topic within their engineering mathematics modules, also offered the students the opportunity to explore the knowledge they had acquired away from lectures. They allowed the students to explore different contexts in which to use their mathematical knowledge.

**Rules**

Like all the other lecturers, Kappa was bound by the rules of the organisation regarding the use of assessment, including safeguarding against the effects of plagiarism. Although she was compelled to act upon plagiarism incidents, she was intent on avoiding such instances by ensuring that the students could not collaborate during summative assessments. To that end, she invigilated the CAA summative tests.

Invigilating the test also provided benefits in terms of the trust Kappa could place in the scores students gained. When Kappa was asked why she invigilated the summative test, she said, “How do you know who’s done it? How do you know that they’ve done it on their own? How do you know if they’ve copied from somebody else or from the book? I invigilate because I want to see what each student working on their own can achieve.”

She was also a member of a wider mathematics education research community, which had an effect on her identity as a lecturer and educator, the object of her activity and what she considered to be good assessment practice. The principles of this good assessment practice manifested as a set of rules by which she was obligated to follow in her assessment regime, which included:

- the need to assess students on their conceptual understanding as well as their procedural capabilities;
- to monitor her students’ progress, and allow students to do the same;
- to offer regular and timely feedback.
The influence of the rules by which she abided had a clear influence on the object of her activity. These rules were beyond the expectations that were placed on her by the department. She had noted particularly the need to test the students on their conceptual knowledge during the semester; she found that the CAA system was not capable of testing students to the extent she required.

Since the tool satisfied many of her aims and brought her closer to the object of the activity, the contradiction was not sufficient to warrant abandoning the tool. The CAA system was also considered a part of the learning regime among the engineering department, since it operated as an adjunct to the workbooks.

**H.1.5 Contradictions**

There were two main sites contradiction that Kappa experienced in the course of assessing her students: a primary contradiction that formed between the identities she had formed within her communities, and a secondary contradiction between the sole use of the CAA system and the object of her activity (Figure H.1).

Kappa developed a leading identity as an assessor, which incorporated aspects of each of the lecturer, researcher and expert identities that informed her practice. One effect of complementing these identities was that that Kappa remained critically reflective with her practice and was perhaps the least influenced of all the interviewed lecturers by learning experiences at school and university. Gustin (1985) suggested that mathematics experts might hold beliefs influenced by these early experiences, but Kappa was a counter-example.

However, there emerged conflicts between identities. Kappa was keen to develop her assessment practice by incorporating technology — using geometry software and having a weekly session for her students in a computer laboratory, for example — but this had an impact on the time she spent on her teaching. Each of the activities associated with these identities demanded much of her time, and she had to deal with the conflicts arising between her identities to manage the time she spent on these activities.

Inevitably, she felt that she could not develop her practice to exactly how she would want to have it had she had the time to do so. She acknowledged that she had to make compromises between her ideals, which were motivated by her identities. Nonetheless, she was satisfied with her practice and it had reached a stable state.

Reaching this stable state relates to the notion of the *the self as a leading activity*, since Kappa exemplified the development of a “real-life activity that most explicitly positions individuals to meaningfully contribute to the ongoing social collaborative practices in
the world” (Stetsenko & Arievitch, 2004, p. 493). That is, Kappa used her experience as a lecturer, mathematics pedagogy researcher and mathematics education expert to develop an identity as an assessor, which in turn allowed her to advise others on their assessment practice.

Kappa faced a secondary contradiction between the object of the activity and the CAA system tool that she was using. As noted previously, Kappa’s interpretation of the rules influenced the object of the activity. Like some of the other lecturers in the study, this rendered the CAA system insufficient for achieving for satisfying the object of the activity.

In Kappa’s case, the need to assess students on their conceptual understanding was not satisfied by the CAA system. Although she was not keen to add conceptually challenging questions to the CAA, Kappa noted that one of the barriers that she faced when using this system was that she could not design her own questions for the system. This meant that she relied on the existing question bank. She also felt that the CAA tests “encourage a more instrumental way of seeing mathematics” that reduces mathematics to a set of procedures to follow.

She addressed this contradiction by introducing group project work. In doing so, she was able to challenge the students’ conceptual understanding and to gain experience of new contexts. Although this was a recent addition to the course, Kappa was content that the projects had served the purpose for which they were introduced. That is, they resolved the contradiction that arose between the object of her activity and the tool.

The introduction of group project work created a new contradiction: it proved to be a burden on her time, which the CAA system promised to save. However, Kappa was satisfied that the time saved by offering CAA tests mitigated against the additional time expenditure for marking the projects.

H.1.6 Effectiveness

For Kappa, the CAA system was effective for the testing of students’ individual work on procedures and it satisfied her aims to assess her students quickly and obtain reliable indicators of their performances. Her CAA practice had developed over several years, so at the time of the interview, her practice using the CAA system had become stable.

She reported that the last major change to her practice was to reduce the number of CAA tests from four to two over the course of the semester, and to make the tests twice as long to accommodate the testing of the material. She reasoned, “With the four CAA
Appendix

Assess students quickly on concepts

Kappa

Assess students quickly on procedures
Invigilated CAA tests
Students’ CAA scores
Students assessed on procedures quickly

Figure H.2: The assessment is effective for Kappa’s goal to assess students quickly on their procedural knowledge, but it is not effective for testing conceptual knowledge.

Tools
CAA practice tests
Paper test

Subject
Lambda

Object
Assessing his students

Rules
Departmental rules
Experiences of assessment internationally

Community
Mathematics department
Engineering department
Other lecturers

Division of labour
CAA offers feedback to students

Outcome
Reduced student context dependence
Differentiating between students
Stable assessment practice

Figure H.3: Lambda’s activity involves the use of paper tests as well as the CAA practice tests to assess his students. He placed additional rules for assessment on himself, taking into consideration his experiences and beliefs of best practice.

tests that I had, many students were finishing very quickly. So, being given twice the number of questions didn’t seem unreasonable.”

While CAA remained effective for Kappa’s goals to assess students’ procedural ability, she could not see a use of CAA as a means to accomplish her goal to assess students’ conceptual understanding (Figure H.2) and, this, the CAA system was not effective for this aim. Although this did not cause problems for Kappa — she was satisfied in using other assessment types for this aim — it served to illustrate one of the limitations of the CAA system.
H.2 Lambda

H.2.1 Subject

Lambda was an experienced lecturer in engineering mathematics, having taught in several different institutions both in the UK and further afield in this area.

His experiences of working with students in different countries gave him insight into the diversity of challenges that students face on an engineering programme. While he demanded much from his students and was strict with assessment, he was confident that his teaching and assessment practices enabled him to understand how well his students were performing.

He applied this experience often as a tutor in the mathematics support centre, providing one-to-one support to students with questions in mathematics.

H.2.2 Object

The object for Lambda was to provide assessment that tested every student. The goals he wished to achieve in assessment that are directed towards this object were numerous.

- He wanted students to apply themselves without being dependent on the context. He was keen that students felt that there was always more to learn and could extend their knowledge further.

- He wanted to gain an insight into his students: he wished to know what topics they were good at and to be able to differentiate between the most and least able.

- He wanted students to practise between lectures; he was concerned that, in the absence of mandatory assessment, many students would not apply themselves between lectures and tutorials and, consequently, they would also forget the lecture content.

H.2.3 Tools

Lambda used CAA with his students in Engineering Department F. Lambda allowed students to access the CAA practice tests but conducted the summative test on paper in invigilated conditions, using questions he had written himself.

He was keen for students to do some work between lectures and providing an assessment encouraged students to do so. He also wanted to encourage students to practise using
the procedures they had learned. The CAA practice tests offered him the tool to provide such a facility. However, he was concerned that using the CAA summative test questions would not provide sufficient challenge and would foster a dependence on certain contexts: “Some of them just repeat and repeat [the questions] and it becomes computerised. If you give them the same question with simple changes, they cannot do it.”

Lambda felt that the CAA summative tests did not offer enough questions to be able to differentiate between students. He believed that the most able students were not sufficiently tested and should be tested on material beyond the scope of the practice tests:

“The reason why is because I looked at some of these tests myself and I can see that the questions are mostly repeated and are simple, and therefore, what I want, is I want to differentiate between someone who could do a normal task and someone could do something more difficult.”

While Lambda felt that the CAA questions were limited and restricting, he felt there was potential to improve the system so that it was better able to achieve his aims for assessment. To improve the questions, he said, “I think what you really need is to set some question with some twists...what you need is that you really need to impose that factor of making the students think.” However, time was a limiting factor, though he would be happy to develop such questions if he was asked.

“It depends whether I have the time or not. That’s something that depends. But I’m quite happy, for example, if I was asked about some questions to contribute. I’m quite happy to do that.”

In addition to the practice tests, Lambda provided students with some optional tutorial sheets that he produced; however he was unsure whether students attempted them: “The reason I give them these practice tests is so at least they have an idea of what they can expect in the paper test and they should really practise. Sometimes, I give them extra tutorial sheets to do outside [of the practice tests], but I don’t know whether they do these or not.”

H.2.4 Community

Lambda felt that his commitments as a lecturer were twofold: he had a commitment towards his students to give them the knowledge and tools they required for their learning
and onward success; and he also had a commitment to his colleagues to provide rigorous and fair assessment to students that offered an accurate portrayal of his students’ accomplishments.

He recounted a discussion he had had about CAA with a colleague. In the following anecdote, he shared his beliefs that students using CAA were becoming context dependent, which affected their success in the exam.

“One of the lecturers some time ago was complaining that the students didn’t do so well on the examination. And I pointed out to him, ‘I think you really need to give them some written tests. The reason being you can prepare them for the final examination.’ Because sitting a test is different from doing a practice test on a computer. So, that’s the only time I’ve come across such a discussion.”

Division of labour

Lambda was part of a group of lecturers that was employed to teach mathematics to engineering students. While this group was not strictly part of the mathematics department, this distinction did not appear to create any contradictions.

H.2.5 Contradictions

He was wary of the CAA summative tests: aside from the context-dependency that he was keen to avoid, he acknowledged that the current system was not actively maintained to update or improve the question bank: “I think we stopped really putting resources in this for a long time. As I recall, I’ve been here now for six years, and I haven’t seen anyone really doing any work on this.”

He had also seen capable students that were willing to abuse the system: “I was seeing this students, who was doing his test on the computer... He was using Maple [a mathematics software package] at the same time and just wrote the answer [that Maple provided]. So how are you really going to check that? ... He knows or she knows Mathematica [another mathematics software package] or Maple, and you put all the questions and could really do it in a second, put the answer and get a full mark.”

Lambda also believed that retaining a human element to assessment was important; he felt it was necessary to be able to offer personal feedback in some circumstances, particularly when the student has made contact with the lecturer to gain more insight.
into the mark he or she had received. He believed that the CAA system was a barrier to human contact: he felt that students were reluctant to approach him for questions about the CAA tests, but were otherwise willing when they had questions about other assessments.

Since the paper tests were invigilated, Lambda was encouraging of students to collaborate when using the practice tests. He understood that, despite his encouragement, some students may be more intimidated to ask him questions about the work than they would be asking their peers. He added that, “I’ve never had any students who approached me to discuss the practice test with me at all over the last five years.” Students had only contacted him to request access to the practice tests in preparation for the exam, which he permitted.

For the paper-based summative test, he was keen to avoid using similar questions to those in the CAA practice tests, saying, “If I use the questions which students have already seen in the practice tests, I’m not doing anything else: they [the students] have an image of the type of question they’re going to expect and, therefore, they are not going to do any extra work apart from this [the practice tests]. But with a paper test, they know they really need to take the extra mile, to make sure they are well-prepared for the test.”

Therefore, not only was Lambda keen to avoid context-dependency, he also wished to convey the importance for students to take extra care and consideration when preparing for a test. This is further evident from the following quote in which Lambda recalls advising a colleague on CAA practice:

“One of the lecturers, some time ago, was complaining that the students didn’t do so well on the examination and I pointed out to him, ‘I think you really need to give them some written tests: the reason being you can prepare them for the final examination.’ Because sitting a test is different from doing a practice test on the computer.”

H.2.6 Effectiveness

When asked about what he believed an effective assessment should be, Lambda described questions that tested students’ abilities to apply their procedural and conceptual knowledge to unfamiliar and practical contexts:

“I think ‘effective’ [is] in the sense that when you set a question, you need to really see whether the students understood the basics of the topic, and how they’re going to apply that for a practical problem. And then...if you could
In this quote, Lambda expressed a belief that an effective assessment should be effective for students, too. It is also clear that his method of writing paper tests in lieu of the CAA summative tests made the assessment effective, in his view. During the interview, Lambda displayed examples of the question he had used in these tests to the interviewer; these questions referred to practical, engineering contexts and some questions invited students to call upon various topics of their previous lectures to solve the problems.

For Lambda, the assessment regime he had employed was effective for his aims. The CAA practice tests were a necessary part of this regime — it allowed students to study and practise at their own pace — but the summative test could neither provide the challenge nor the ability to differentiate students on their ability and understanding. Thus the summative test was not effective, since they would not satisfy Lambda’s aims for assessment.

Although Lambda did not express any desire in the interview to change the CAA summative tests to suit his aims, it would appear that Lambda desired more complex questioning and more of an engineering context in the summative test before he would consider using it. However, Lambda was also more keen to retain a human element to the assessment, which is at odds with Lambda’s perceptions of the CAA system.
H.3 Mu

H.3.1 Subject

Mu was an experienced lecturer whose research was in mathematics education and, in particular, cognition at undergraduate level. She had a national and international reputation for her research contributions in this field and she was well-regarded as a lecturer among her students.

She developed identities as a mathematics education researcher and as a mathematics lecturer. There was some influence between these two identities: her research helped to inform her teaching practice and her observations while teaching helped generate hypotheses and ideas for her research.

H.3.2 Object

The object of Mu’s assessment activity was to assess her students in their mathematical knowledge. She partitioned this object into several goals that related to the level of understanding and the quantity of knowledge:

“Ideally, in maths [assessment] you want to do two things, I suppose. You want to test whether people know the stuff — so whether they know the definitions, whether they can do the standard things, whether they’ve shown
some kind of mastery over the topic, which I tend to think of as sort of an up-to-2:1 level of an answer... But you also want to test, ‘Can you do a bit more than that? Have you understood things sufficiently well that when you’re presented with something that’s a bit less predictable, you can put some things together, work out how to apply something things that you know and do that?’

When she asked whether the CAA system satisfied her definition for ‘effective assessment’, she responded, “Well, in that sense, no, I suppose. If you just asked about a literal match to what I’ve just said, no, it doesn’t… [Though] to some extent, yeah, because you’re testing, ‘can they do the standard stuff?’ ” Therefore, Mu did not believe that CAA was suitable as a replacement for all assessment, since it did not achieve what she wanted to achieve in assessment.

H.3.3 Tools

She used the CAA system to assess students in the Mathematics Department. Mu chiefly used CAA since it was firmly established on the course when she started teaching it; she continued the practice since it provided a means to test and offer feedback to a large number of students in an efficient manner.

The historicity of CAA being firmly established in the course was the overbearing reason for Mu to continue using it: it had been used successfully by a previous lecturer of this module and Mu had not been convinced to stop this practice. Other lecturers cited established practice as the primary motivation for using CAA on their courses.

Mu was not in favour of CAA being the only means of assessment: she was cautious of missing out on assessing other aspects of knowledge that the current system could not assess. She felt that the CAA system was not particularly well-suited to assessing conceptual understanding. Nonetheless, Mu noted that the system was good for assessing procedures, thus she continued to use the system for this purpose. She felt that exams were the appropriate opportunity to test students beyond just the procedures, and so she felt that she had reached a stable practice that assesses students comprehensively in a range of skills:

“I’m quite satisfied with that [CAA] system for part of the credit, because it does mean that I can ask proper questions and send people away to properly think about things, while still testing individuals in that sort of way. I wouldn’t want to replace final exams with that sort of thing, because I think
they [the exams] do force you to really think, really try, and not just think things through, but get to a point where you know the stuff sufficiently well that you can remember lots and lots and lots of it [the material]...I tend to believe that although people have all these negative things about summative exams, in order to do well at it, you have to able to do that harder stuff, I think. You really have to be able to get it together, and it’s not just a case of memorising.”

She felt that the CAA system was particularly weak for the content of her course, in which students must memorise and apply definitions:

“There is a very good sense in which it’s not really to my benefit to spend however-many hours it takes me marking lots of people’s definitions [relating to the module], for instance. If I could get a machine to do that, that would be great, because I would want to see it written down, as logical changes are important and all that kind of stuff. . . . It would be better if that could be done in a way that didn’t involve me spending so much time because it’s very standard.”

H.3.4 Community

Mu was a member of two key communities that had an influence on her identity and the activity of assessing her students: the mathematics education research community and the mathematics department at the studied institution.

Being a member of the former community made Mu more aware of developments that had happened in CAA elsewhere:

“I have been to talks where, for instance, [someone] did stuff [at another institution], talking about questions that are more — I want to say ‘adaptive’ again, because that word’s in my head — but ones where you get potentially better feedback, or ones where you get hints, or ones where the numbers are properly randomised, rather than just a choice of ten questions. And that sort of thing is very appealing. I gather that Maple [mathematical software] and various other systems have various different types of these questions.”

Because of this awareness of the other tools available, Mu believed that it was possible that the system could be changed soon: “it may be that we’re all using things that are a bit better in the next few years, anyway.”
The prospect of change and the difficulty in making changes to the current system meant that she did not spend time making changes. With a new system, where changes could be made more easily, she could envisage spending some time making improvements:

“If it became easy. Which is not to say that I don’t want to continue making things better. It’s just that there comes a point where, for small things like this, ‘Am I going to do that? Or am I going to improve my lectures? Or am I going to say that everything’s good enough and do more work on my research?’ You know, there’s always that trade-off.”

Here, Mu echoed Kappa’s concerns that there was a tension between teaching and research identities: time spent tending to an activity relating to one identity meant less time available for activities of the other. She explained in this passage that there came a point where she believed that her practice was “good enough” and that her time was better spent on research, rather than on making small changes.

However, Mu was clear that her practice was not always influenced by research from the mathematics education community because sometimes the advice from research is conflicting:

“As I said before, I could re-design things so that I thought they were better. But that would take a lot of time. It’s not clear what the gain would be. And it’s not clear that I have any evidence about what would be better. Most of my teaching decisions — even despite the fact that I work in mathematics education — most of my actual teaching decisions are not evidence-based to the extent that it would be ‘ideal’. Partly, of course, it’s always difficult to work out what would be ideal, because the evidence is often so contradictory.”

Mu explained the role of the department in the decisions that she and other lecturers in the department made for administering their courses:

“That’s one good thing about being an academic. Basically, if you want to change something on your course, and you’ve got a reasonably sensible reason for doing it, you can just do it. Sometimes, you have to put things through committees and stuff if it’s a big change.”

While she enjoyed these freedoms in her teaching, there was a responsibility to ensure that students were not inconvenienced by these choices. Mu explained that there should be some continuity between two semesters of a module when there is a change of lecturer:
“I guess it’s important in this two-semester course to be somewhat consistent from one semester to the next in how these things are set up, so that the students are not confused.”

**Division of labour**

Mu initially found it time-consuming to set up the CAA system for her teaching and assessment; however, after this initial outlay she found it more straightforward:

“It took me a while to decide first time what questions I wanted to use and stuff, but it’s now pretty easy, because I just use the same ones again. It was used by the previous person. So, the system is all set up, so I don’t have to do any of that. Roughly when things happen, I just have to tell someone when I want it. Everything happens without me having to do anything, so that’s good.”

The assistance provided by support staff made the job of assessing students easier, and this division of labour was gratefully acknowledged by Mu and other lecturers.

Mu felt that students had a significant part to play in their own learning through assessment. She felt that it was not the role of assessment to provide students with the motivation to practise — “those who are motivated will do it; those who are not will probably leave it to the last minute and not get the full benefit out of it” — but the assessment should be adequate enough to support those that are willing to learn:

“There’s a sense in which I sometimes worry...because some students, at least, are being a little bit obsessive about tests and spending, probably, disproportionately too much time on that, compared with other things. ...I’d want them to practise. I wouldn’t want them to become obsessive about getting ten out of ten, because that’s probably not a good use of their time.”

**Rules**

It is not transparent from the interview where her rules for how assessment should be conducted originated from. However, her identities as a mathematics education researcher and as a mathematics lecturer within a mathematics department may have been sources for these rules.
For each of these identities, there was a community in which Mu was a member. As noted with Kappa, there would be expectations placed upon members of the mathematics education research community to be aware of new research and updated practice. Furthermore, the mathematics department and the wider university teaching community place expectations on lecturers in terms of assessment and teaching practice. There may also be influences from her experiences as a learner (Gustin, 1985).

An example of how these rules affected her practice was to be found in the feedback she offered to her students, and in CAA in particular. She noted that it was not feasible to achieve her ideals for feedback with so many students in the cohort. When asked what would be better feedback, she said:

“Me sitting in a room with each individual, if that’s the level that you really want for that sort of feedback. And obviously you can’t achieve that in that sort of class.”

Mu acknowledged that this was not a realistic expectation. However, she felt that the CAA feedback was not as good as it ought to have been:

“It’s not good quality feedback. I mean, it’s model answers, which is not terrible: it’s not just, ‘Yes; no; you were right, or wrong,’ which would be worse. But it’s a static thing. It doesn’t respond to the student. It doesn’t give them a chance to have another go.”

Because students were not getting individual feedback that responds to their work, Mu felt that the feedback was “poor quality”. However, Mu balanced what the CAA system offered her students against this shortcoming. For Mu, the exercises provided by the CAA system were an opportunity for the students to practise, which they should be doing anyway:

“For routine exercises, the students should be doing those anyway. It might be good if they had access to that sort of thing all the time, so that they could do the test on their own and study, rather than it necessarily being part of an assessment scheme.”

Mu also discussed the rules she communicated to her students in terms of how they should work on the CAA tests. She was keen for students to collaborate during the practice tests, but discouraged collaboration during the summative test. She did not
communicate these wishes to her students, however. In the practice tests, she realised that some students were more inclined to collaborate than others and did not want to impose collaboration on students that wished to work alone:

“I haven’t encouraged them in any meaningful way, and I don’t know that I would actively encourage that. I don’t care either way, really. Some people tend to work on their own and some people don’t. I don’t encourage it, I suppose. Maybe I should. I mean, I want students to be talking about maths generally. If they happen to be talking about maths for the test, that’s fine with me.”

She believed that discouraging students to collaborate in the summative test would be a futile exercise and did not want to discourage students from seeking help if they were struggling:

“I would want people to find things they don’t understand and discuss it with other people, go through their lecture notes together, work on problems together. Not necessarily a lot, but to some extent, I think it’s good for them to try and speak about mathematics in a coherent way and practise that.”

…I don’t think there’s any point [in discouraging collaboration in summative tests], really. If they’re going to do that, then they’re going to do it. Since I don’t actually have any actual means of controlling them, then me even saying that — maybe I should — but there doesn’t seem to be a great deal of point. If somebody is determined to try to get someone else to do the test for them, then there is nothing I can do. I would never know, and I would not be able to stop them.”

Although there is a rule in place within the department that prohibits students from plagiarising, Mu found that it was impossible to police in practice. However, to mitigate against this concern, Mu was satisfied that the contribution the tests make towards the overall module mark was sufficiently small: “I’m not sufficiently worried about it to really make my own life and their much more difficult by starting to run it as an invigilated test.”

H.3.5 Contradictions

Mu felt that the system was not able to fulfil all her aims for assessment. She noted other contradictions: the system offered immediate feedback that students could respond
to instantly, but Mu was unsure of the quality of the feedback itself. She said, “The immediate feedback thing — yeah, it depends on the stuff. I don’t use it because of that. I mean, it’s good that they’re getting some sort of feedback, but...the feedback quality is not that high, really...I don’t think it’s good-quality feedback in the sense of being individual, or being able to give hints, or anything like you would do if you were in a room with a person, which is a completely different level of helping them to think”.

This compares with the wishes of the Lambda, who was keen to retain a ‘human’ aspect to assessment. Lambda felt it was important to preserve the possibility of communicating with students individually if they experience issues with their work. When Mu was asked about the interaction she had with students, with respect to the CAA tests, she noted they were technological or functional in nature, rather than related to the mathematics. She sometimes responded to queries while in the mathematics support centre: “someone with want more detailed feedback or something like that, or they’ll want to understand the feedback better...but by and large, I don’t get anything other than procedural queries.”

There is a further contradiction with respect to her aim to encourage students to practise. While acknowledging that it was broadly a good feature of the CAA system, Mu was also concerned that students spend too much time and effort on these tests: “There’s a sense in which I sometimes worry about that, because some students, at least, are being a little bit obsessive about tests and spending, probably, disproportionately too much time on that, compare with other things.” She added “I’d want them to practise. I wouldn’t want them to become obsessive about getting ten out of ten, because that’s probably not a good use of their time.”

Despite these contradictions, Mu continued to use the CAA system with her students and acted within the constraints and contradictions. With such a large cohort — “it’s 200 people” — Mu was willing to provide an assessment that would fulfil her simpler aims and use other assessments to achieve her other aims.

The size of the cohort meant that invigilating the summative test would require excessive effort to achieve. On balance, Mu believed that the risk of students cheating was not significant enough to warrant a reconsideration of using the system: “I’m not so worried about them doing it inappropriately to be honest. It’s a small part of the course. I don’t want them to be quite cheating, but on the other hand, people who are going to engage seriously with this will do better later anyway.”

Mu was happy for students to talk about the mathematics and the questions they were doing. With regard to the students, Mu’s aims for the assessment was to encourage students to “[think] about it” and to “keep them working”. She was keen for her students
to routinely discuss the mathematics they were using: “I haven’t encouraged them in any meaningful way, and I don’t know that I would actively encourage [collaboration]. I don’t care either way, really. Some people tend to work on their own, and some people don’t.”

H.3.6 Effectiveness

The CAA system accommodated many of Mu’s simpler aims for assessment, but she acknowledged that it was not particularly suitable for her aims to test her students to a deeper level. She believed other colleagues felt the same and that there was a wider frustration with the system: “I think that the system that is there — everyone is somewhat dissatisfied with it.” However, she conceded, “But then, everyone is somewhat dissatisfied with whatever the system is. It’s like no-one ever likes everybody’s else’s textbook.”

Nonetheless, Mu conveyed mixed feelings with regard to the system, acknowledging that it does satisfy some goals but that there was a desire for it to be able to do more: “I’m not excited by it, but I’m satisfied with it, I suppose. Am I slightly dissatisfied? I don’t know. I mean, there’s a sense in which, yeah, I could be.” As for the desire for the CAA system to assess conceptual understanding, Mu said, “It’s hard [for CAA] to test for deeper understanding of mathematical concepts, I think. But it always is, so I don’t know if that’s a particular fault of these kinds of systems.”

Mu was satisfied that the CAA system was sufficient for four main goals. There is a suggestion in what Mu says about the ways in which students might collaborate that it is difficult to ascertain exactly how students act when undertaking the summative tests, with some suspicion that cheating might have taken place without a realistic prospect
of it being identified. Without knowing for certain how the students acted, it would be difficult to ascertain whether some of Mu’s aims for the assessment had been achieved.

Nonetheless, it remains that the CAA system was effective for Mu until an assessment of students’ conceptual understanding was required. For this, Mu employed different forms of assessment.

**H.4 Nu**

**H.4.1 Subject**

Nu was a lecturer in the mathematics department whose research background was in mathematics. He had formed close links with members of the mathematics education department at the studied institution and was familiar with and interested in sharing good teaching and assessment practice.

His identity manifests in his commitment to the students on his course. He keenly responds to module feedback and offers opportunities for individual feedback. This is reflected in his belief that, “Ideally, [the students] should have a lecturer watching them do the work and then comment on what they have seen.” He was a strong advocate of small group tutorials that offered these opportunities.
Appendix 242

H.4.2 Object

Nu’s aims for assessment were to encourage students to practise, to expand and develop students’ views of what it means to “do mathematics”, and to encourage students to “learn certain things”.

He had a desire for students to experience mathematics to a deeper extent than is strictly necessary for the course content. In order to achieve that, he offered a variety of different assessment techniques, including an assignment in which the students had to “find and correct five mathematical mistakes” in a given piece of text. He felt that students would work only on assessment that they are given credit for, so he adopted assessment as the opportunity to expand students’ experiences of what it means to “do mathematics”.

H.4.3 Tools

Nu started using the CAA system when he started teaching the module. It was already established on the course and he continued that practice. When asked why he used CAA on his course, he said, “Basically, because I’ve inherited it. That’s the first answer. But also because it makes students practise.”

Like Mu, Nu felt that CAA encouraged students to practise but not always in the manner that was intended. In particular, there were students that would put excessive effort into CAA tests: “There are students who do it dozens and dozens of times — not very many — but that happens.” Nu monitored how his students used the practice tests; he found most students use the system responsibly but some patterns raised suspicions:

“Most students will do it once, twice, three times, but it still comes out to...I don’t know. For me, it was a surprisingly high average of how many times students do these tests...Sometimes, I look at the times [durations] that a student has taken [to do the tests], and I’m a little suspicious about how they could have done it. Certainly, for the numerical questions, I do suspect that a number of those are answered by Maple.”

It was a particular strength of the CAA system that Nu was able to monitor his students’ working habits, something that he could not readily do with tutorial sheets:

“If I set tutorial problems, I don’t know what’s going to happen with them. A number of students are going to do them, and a number of students don’t. Basically, I have no way of knowing. With the CAA tests, I know that a lot of students do them, and for that reason, I think it’s worth doing.”
Nu was one of the few lecturers that helped to develop the tool. He wished to test his students more deeply than the existing question bank was capable of doing. Consequently, he wrote some new questions for the system with the intention of assessing some aspect of the students' conceptual awareness. However, he felt that the limitations of the system meant that he could not pursue this approach of testing conceptual understanding further:

“But that’s as far as I’ve been able to push it, in terms of getting towards testing students’ conceptual understanding. Whereas, of course, in terms of being able to do calculations, that works very well.”

H.4.4 Community

As well as being a member of the mathematics department, he also sometimes participated seminars that were run by the mathematics education department, including a series of talks addressed towards sharing and discussing mathematics pedagogy, assessment and issues in learning. Although his research was not in mathematics education, he was committed to learning more about lecturing and was willing to engage with others to reflect upon and develop his practice.

As such, Nu was a member of staff committees in the mathematics department where issues like assessment were discussed. This helped him to develop an awareness of how CAA is used across the department and the practices that other lecturers had developed and used.

With regard to CAA, he had discussions with the teaching coordinator and to the CAA administrator about his practice and the developments he had introduced. This helped him to become more aware of how his practice influenced the way his students approached CAA. For example:

“A few years back, the teaching coordinator asked me to have a look at how many students actually do that [the practice tests], and it turns out that they do this quite a bit. There are students who do it dozens and dozens of times — not very many, but that happens. Most students will do it once, twice, three times, but it still comes to . . . a surprisingly high average of how many times students do these [practice] tests.”
Appendix

Division of labour

In order to deliver CAA to his students, Nu used the services of the CAA administrator to set the tests and to ensure that the system’s settings are as desired. Nu noted occasions where he desired changes to be made to questions or to the system; he regarded the administrator as reliable and able to make such changes:

“I don’t know what kind of files [the CAA administrator] still has for that. But if I tell him, ‘This problem, this particular version, needs a correction’, then he can do that. . . . But his answer has been, ‘It’s not a problem at all. I can fix that.’ And he does that.”

He realised that his mathematics module was running at the same time as other mathematics modules. He talked with lecturers of these modules as a matter of course and was aware that practices were different and that there was a case for consistency.

“We have discussed computer-aided assessment in [a committee] a number of times and, in particular, we have discussed there to what extent students might get help from other or get help from Maple [mathematical software].

...I know, in [a concurrent mathematics module], or in some other modules, the practice test goes away when the real test starts. I have considered that. I can’t see a compelling reason to do that. But the students keep asking me about when the practice tests will go away, probably because they do go away in [the other module]. I think there probably is something to be said for doing the same thing in both modules.”

Rules

Although Nu did not explicitly tell students that they are not allowed to collaborate during the CAA summative test, he explained that he had put the departmental and university policy on assessment and plagiarism on the virtual learning environment for students to see and adhere to. When it came to CAA, Nu felt no need to enforce this policy, since: “for 2.5%, I just don’t think it’s worth putting up a major police operation to find out what students actually do.”

He felt that the policy was open to interpretation, however. Nu explained that the policy related to written coursework submissions and that “the school coursework policy is that
students are allowed to discuss the problems, but everybody must write up his or her own solution. Now, for computer-based tests, of course, there isn’t a lot of writing up to do.”

H.4.5 Contradictions

Although he was suspicious of this potential cheating, he remained unconcerned: “That doesn’t worry me because these the basic calculations things that students will have to master anyway. And they do.” He added, “The weighting of these courseworks is very low. . . . Each of these courseworks is 2.5%. Because it’s so low, I’m not really worried where students take their answers from, as long as a sufficient number of students puts in the work for to prepare for the test and learn from what’s there. If that’s happened, as far as I’m concerned, the test has served its purpose.”

Invigilation would, perhaps, have solved Nu’s problems. But, as in Mu’s case, the cohort was too large to implement invigilation in a straightforward manner: “I think it’s more hassle than it’s worth, really, for a test of that size to arrange for invigilation. You know, the groups are large. . . . arranging invigilation for that would be a huge task. And I don’t think that what we gain from it is worth the effort.”

Like many of the other lecturers in the study, Nu was unsure quite how students approached these assessments. When discussing collaboration, Nu conceded that he was unaware how students acted during CAA testing and that his ideas of how students may collaborate or cheat were conjecture: “They can, of course, collaborate in the sense that they can have a group of five standing around while they are taking the coursework test and they they could move onto the next person. I have no way of knowing whether that happens. I think that, clearly, would be a violation of the coursework policy.”

Nu noted that CAA testing produced a grey-area in terms of the coursework policy. While he believed that students gathering in that way would be cheating, some types of collaboration may be considered cheating by the lecturer, but not in terms of the coursework policy: “The School coursework policy is that students are allowed to discuss the problems, but everybody must write up his or her own solution. Now, for the computer-based tests, of course, there isn’t a lot of writing up to do.”

Nu recalled that there have been multiple instances in which such issues were discussed at lecturer committees: “We have discussed these to what extent students might get help from others or get help from Maple. Not everybody takes the relaxed attitude about the computer-based [tests] that I do.” In one such meeting, Nu participated in a discussion in which another lecturer had switched from a non-invigilated test to an invigilated one:
“I do remember that we looked at the outcome, it was very striking that marks had down noticeably compared to the non-invigilated tests. But we came to the conclusion that we weren’t really sure if that was due to the invigilation or whether it was due to the fact that they invigilated tests were closed-book.”

Nu’s recollections revealed that there were more widespread concerns about CAA testing between colleagues in the mathematics department. It is also apparent from both these lecturers’ responses and Nu’s recollections of these meetings that there was no consensus on the best practice and that lecturers’ practices varied. Furthermore, the extent to which the lecturers were fazed by the problems that CAA presented differed: the lecturer in Nu’s anecdote appeared to be concerned enough to change his practice, whereas the lecturers in this study appeared to have established their practice and were sufficiently content to maintain it.

A further point of note to consider is that the lecturers appeared to be reassured by the lower marks that students were gaining from CAA as a result of the change of practice. Perhaps this is an indication that it gave those lecturers more confidence in the system that it would satisfy their aims. In particular, a lecturer that aimed to distinguish between students based on their ability may have wanted to see greater variability in the scores.

One conclusion that can be drawn from this is that there are lecturers for whom the CAA system did not sufficiently satisfy their aims, which resulted in contradictions: the lecturers were keen to continue using the system, yet they felt that the system was not offering them what they desire to achieve at that time. For one lecturer, this proved sufficient motivation to change his practice and to report his findings to the teaching committee, where his practice would be shared and discussed. In this instance, the outcome was that other lecturers were interested in the results, but were not convinced by the cause of the change in marks. This further highlights the problem that lecturers did not know with certainty where students obtained the answers to the CAA questions.

Nu echoed the views of Lambda, insomuch that he believed that the students were being presented with familiar questions in the CAA tests. Lambda was concerned that students would become context-dependent and consequently addressed his concern by writing paper tests that introduced new contexts. Nu said, “Because I’m more or less bound by low-level computational things, I end up asking the questions that students expect to be asked, even though I’d much rather try to move away from that expectation and change their view of what it means to do mathematics.” However, he added, “But for that, there are other assignments in which I can do that.”
In Nu’s case, this did not present as much of a problem. His primary aims were to encourage the students to practise using their newly-acquired mathematical knowledge between lectures and to consolidate their learning. An important part of this process was to give students timely feedback, though Nu was not entirely satisfied with the feedback that students received. He believed that the feedback was too static and did not respond to the students’ answers: “There is a screen of feedback for each question that comes up if a student puts in a wrong answer. So, that’s not the level of detailed feedback you might want to give, but that’s what there is, and that’s what we reasonably can do.”

Nu later expressed his concerns about the feedback in more detail, suggesting that it might not be considered ‘feedback’ at all: “In a sense, it isn’t really feedback because it doesn’t take into account what the student has actually done.” He also believed that the feedback students were given “isn’t much more than a worked example, as you find in the lecture notes.” As well as the feedback the system gave to students, he was the only lecturer to comment on the feedback he received from the CAA data and noted that it was difficult for him to establish how students went wrong in their tests: “All I have to go by...is the one number that the student has put in...I have very little knowledge of what the student has actually done.”

Although the feedback was not to Nu’s preferences, he offered students the opportunity to get more feedback from him if they desired: “I do say to my students in the lecture that that level of feedback is not what I would like them to have, and I encourage them to come to me and ask for more feedback if they want it. That happens very rarely, but two or three out of 250 are going to do it on every test.” Here, Nu acknowledged that there was a consequence to offering additional feedback in that to do so takes time. Since such a small number of students requested this feedback, it was not so much of a burden on his time. Nonetheless, while CAA offered to save lecturers’ time by administering feedback automatically, Nu remained committed to reinvesting his saved time to offer feedback after perceiving inadequacies in the CAA feedback.

There were other problems in the system that caused Nu frustrations. There have been occasions where there have been mistakes in the system — “embarrassingly, even in problems that we actually have run for ten years, long before my arrival here, last year, students did find mistakes.” There had also been problems in administering the tests when Nu wished to limit how often students could attempt the test:

“The test is set up such that students can — in principle — do it as many times as they like, and they get a score recorded the first time they get a score. To some extent, of course, that means that they can choose the questions they are given [by loading a new test without submitting a
score)... In one year, I had asked [the technician] such that [the number of times the test could be loaded] was limited to two.

“It turns out that runs into technical difficulties and I had dozens of complaints... and I had to take care of that... I discussed that with the CAA team and, basically, their answer was, ‘That can’t be, the students are lying.’

I'm not totally convinced of that, because I had, really, many students coming in with a description of that problem.”

This suggested that Nu was not entirely trustful of the system — indeed, he trusted the testimonies of his students more than the system and its administrators, despite being somewhat suspicious of the actions of some of his students. This lack of trust may be due to other points he raised regarding the CAA system, noting its age, the mistakes that have been spotted and the limitations in the type of questions that the system could ask: “The CAA is limited in the way that I can either set multiple choice or I can set a numerical answer question. There are a couple of other possibilities on the system, but if I really want to check students’ deeper understanding... then I need to say, ‘Explain to me why.’ And that I’m not going to get... with the computer-aided assessment.”

It is also evident that Nu was somewhat frustrated by the fact that the changes he had attempted to make to improve his trust in the system had not worked. He had invested time in making changes to the system for this reason, which introduced another contradiction. The CAA system was intended to save lecturers’ time, yet Nu recounted this instance where he has expended time that transpired to be wasted.

Yet Nu expended time on the system in other ways: he was one of few lecturers that that had attempted to write new questions for the system. He used mathematical software to design a sufficient number of similar questions to satisfy the randomisation process. This proved to be a significant burden on time, but allowed him to ask questions aimed towards testing the students’ grasp of the terminology:

“I have put in a couple of questions. For example, there is one question that says, ‘I’m going to give you two statements. One of these statements is true...’ Then I give a pair of statements, which say, ‘Given the matrix A, the nullity of this matrix is 2.’ The other statement says, ‘The null space of the matrix is 2.’ And with that sort of thing, at least I get access to students’ use of the technical terminology.”

Like other lecturers in the study, Nu used the CAA tests alongside other assessment types to address all his aims for assessment. That the system could not be used to ask
every type of question that Nu wished did not post a significant problem since other assessments were more capable of this task: “I can’t ask my students the questions that I’d really want to ask in a [CAA test] coursework, so that’s why we have both.”

H.4.6 Effectiveness

Considering the effectiveness of the CAA system with respect to Nu, there are aspects in which the system excelled but the nature of the feedback and the limitations in the type of questions limit the scope of its effectiveness.

While the system proved effective for encouraging students to practise, Nu remained concerned at their practice habits: in particular, he said, “CAA isn’t necessarily good at making students practise the right things.” The feedback was also not to Nu’s desires, and Nu remained concerned that the system was not completely reliable.

Although Nu was prompted during the interview to recall discussions he had had with other lecturers, he was detailed in his recollection of the discussion that arose from his colleague’s experiment with new practice. Although Nu gave no indication that he had considered changing his own practice, it was clear that he maintained an interest in his colleagues’ experiences. Nu had noted that he shared concerns about the system with other colleagues, but was less concerned by their impact. Nonetheless, Nu was particularly attuned to the contradictions faced by other lecturers in his community; yet while other were changing their practice and reporting their findings, Nu was satisfied with his practice and the steps he had taken to address those contradictions.
Therefore, the fact that the CAA system was no longer effective when Nu intended to test deeper understanding and to offer some more personalised feedback was not a problem, since Nu employed other assessment types. Nu desired improvement in the system to address the shortcomings but did not believe that it was viable with the current system.

H.5 Xi

H.5.1 Subject

Xi was an early-career lecturer in the mathematics department that had experience of teaching in schools and whose research focus was in mathematical pedagogy.

Her research interests were in mathematical pedagogy and, most recently, at higher education level. Before starting her career in research, she had been a teacher of mathematics at secondary level. She acknowledged the importance of all three of these identities — the lecturer, the researcher and the school teacher — when considering her teaching practice with students now: she used phrases in the interview such as, “As a lecturer, . . .”, “From my own research” and, “Having been in schools, . . .”.

Having these identities helped her to relate to her students in ways that some of the other lecturers could not: she was teaching first year students that were adjusting from school to university study. In this transition period, she was aware that students were less likely to display frustration when they were failing to master the content:
“I understand their frustration, having been in schools where pupils complain a lot more visibly and get upset. But these students are adults now, and they’re frustrated in the subject that they want to do well [in].”

Nonetheless, she believed that university should be quite distinct in the nature of learning to what it is in school education. She felt that modules and persistent assessment were somewhat against how she viewed university education:

“I don’t want university education to be regimented. I think it must be creative enough, individual enough and self-motivating enough. It must reward those who have been going out of their way to spend time studying their subject. And if you test too much and straitjacket students into focusing on assessment all the time, producing projects and coursework, I think it’s against the nature of a university education, or what I think it should be.”

H.5.2 Object

Like many of the other lecturers, Xi wanted to achieve many things in assessment. As well establishing the knowledge that her students have acquired, she wanted to offer a means of acquiring motivating and useful feedback. An assessment that offered such feedback would also help to develop their confidence. However, with her time split between research and teaching, she realised that she needed to make time savings: she did not wish to spend too much time on assessment.

H.5.3 Tools

Xi used CAA with her students in Engineering Department G. She operated an invigilated CAA test that followed a period of free access to the practice tests in the same manner as Kappa. Her reason for doing this was from her aim to provide a fair and equitable assessment for her students, since she was concerned that some students would gain advantage from the lack of invigilation: “I could not really let them take it outside, because I could not be sure if it was fair, in the sense of some used materials and some didn’t.”

Like other lecturers, Xi conceded that it was a concern that she would not be able to monitor how students undertook the summative test if it had not been invigilated: “I didn’t want them to take it in the halls of residence or with friends, because you never really know about it then if I’m not there.”
Xi was keen to use CAA with her students to offer them opportunities to practice, address their weaknesses and consult with their peers. Her view was that the practice tests allowed students to take an independent and mature approach to addressing problems:

“A practice test is not just practising getting the right answers, but understanding what the questions are, how to go about them and — if they have a problem — to find out what it is they have a problem with.” This was part of a wider strategy to foster the transition between school and university learning; she desired that students made good use of the resources available and to manage their own learning. She said, “I don’t mind if they ask their teacher, their friends, their parents, anybody; consult books — as long as they find out where they’re going wrong and to improve their own scores.”

H.5.4 Community

Xi seemed to regard the students as part of the community in her activity of assessment, more so than the other lecturers. She discussed the impact of her decisions for assessment on her students more than the other lecturers and this was a constant theme throughout the interview.

She was keen to engage with her students in discussions, noting that, “If you start listening to the students, they start complaining!” That is, Xi was aware of many of the problems that the students in the study presented. One such issue was the nature in which marks were awarded:

“There were complaints each year about students feeling that they didn’t get enough marks for the amount of mathematics that they knew. They were frequently failing on the detail, on the mathematical details and carrying out integration and differentiation and getting no marks, even though they were doing the right thing.

“They knew it was an integration by parts [problem], but were failing to put a factor in here or there, or an angle was missing. They got no marks out of the work.”

Xi appeared to be more attuned to her students in terms of the way they worked on the CAA tests and how they felt towards the system. Of all the lecturers in the study, only Xi had talked to students about these issues. When asked how her students collaborated in the CAA tests, she said, “I sometimes have found out informally when I talk to students
in tutorials. But I wouldn’t know in detail, but they definitely are working together in
different ways. I have heard in round-about ways and through my own research that
students, if they live together in halls in the first year, they tend to be near each other,
room-wise; so then they will start talking about the work then.”

**Division of labour**

Since students were an integral part of the community in Xi’s assessment, she contended
that they had an important role in their own learning and that one purpose of the
first year of university education is for students to acquire a sense of autonomy and
self-direction for their studies:

“I suppose it’s part of the first year. . . . They’re all new to the university and
I think it’s [her advice to students] an attempt to get the message across that
they should be making collaborations, making friends, and discussing work
is ok.”

Although she did not feel that other lecturers had much influence on the way she used
CAA and assessment, Xi continued the practice of the previous lecturer of her module.
While there was an administrator for setting CAA tests, Xi suggested that she selected
the questions herself. Therefore, the CAA system offered a division of the labour of
assessing students, however she might have saved more time had she approached the
administrator.

“There used to be quite a bit of work involved in setting it up, because you
used to go on a computer and select the questions, which were all pre-given.”

While the CAA system handled the questioning, Xi lamented the limitations it had:
“The downside is that I have very little say, really, in what is being tested and how it is
marked.” So while the tool offered to divide some of the labour of assessing students, Xi
felt that it did not do so adequately.

**Rules**

Xi was more keen than other lecturers to encourage collaboration between her students.
This was particularly the case in this first-year module since she believed that acquiring
sufficient knowledge of the material is necessary for the students’ success in future years:
I’ve told them to use whatever means they have, including fellow students, ...because I suppose it’s part of the first year ... so they’re all new to the university, and I think it’s an attempt to get the message across that they should be making collaborations: making friends and discussing work is ok. Some of them arriving here are quite isolated.”

While most of the lecturers had suspicions of cheating, Xi had evidence of students cheating when she had run the course in the past: “In the first one or two years, some students did abuse that [the invigilation] and took the test outside of invigilation. So, I had some attempted cheating.”

H.5.5 Contradictions

This was contrary to how Xi would have marked it herself; she felt that marking in this way was harsh and demotivating: “If I had hand-marked that work, they would have got a higher score, because I would have marked for the correct procedure used and been lenient with some of them on the detail. Now, the CAA test doesn’t do that... In retrospect, with these weaker students they were getting scores of around 40%, which is very low for CAA testing. Usually you expect seventy, eighty, ninety per cent.”

This proved to be an unresolved contradiction for Xi, as she continued to be remorseful as students became upset with their scores. She acknowledged that her students were under significant external pressures to succeed and that the CAA system proved to be somewhat of a hindrance to some students’ aims. She noted that, “it’s just the whole wider context of studying in the UK and in a time of high unemployment, even for graduates. So they need to do well.”

She felt that the system better suited those students that had grasped the material quickly and were able to achieve high scores: “I’ve come to conclude that CAA testing advantages the better students, even though it’s a nice tool to practise questions for the weaker students. They gain a lot of confidence if they’re getting it right, if they’re getting good feedback. When it comes to a one-off coursework assessment, if they’re nervous or if they forget, or they’re missing something here or there, they’re ending up with zero. And so, I came to the conclusion that it benefits the stronger students and not the weaker students.”

She added that, for the stronger students, the feedback was useful for identifying exactly where they went wrong when carrying out a procedure; however, for weaker students who were struggling to carry out the procedures correctly, the feedback was less useful: “The weaker students who would call me over say they did this work afterwards and
Figure H.10: Xi’s main aims were to provide timely yet encouraging feedback to students so that they gain confidence and progress. She found that it did not do so for weaker students.

showed me correct working on two pages, having gained no mark and being quite upset about the fact that they had struggled really hard to get on top of the mathematics.”

This was a significant contradiction that Xi faced: she had particular concern for the weaker students since she felt that these students were the least confident and most likely to be discouraged. To that end, she felt that it would be “wrong” to label the CAA system as a good assessment. She noted that she had used project work as another means to test her students, which perhaps alleviated this contradiction — as it had for other lecturers — however, Xi did not mention this in her interview.

She continued to detail how the system made her “feel bad”:

Initially, I was quite happy doing it, because it was very little work. It was easy. Coming in [to teach this module], it was prescribed — just pick a few questions. It took me half an hour. No marking. Four tests that way. but when I get it back from the students, I feel bad because I understand their frustration, having been in schools where pupils complain a lot more visibly and get upset. But these students are adults now, and they’re frustrated in the subject that they want to do well.”

H.5.6 Effectiveness

While the CAA system satisfied her aims for assessment with respect to her workload, Xi felt that it was not adequate for her aims with respect to encouragement and progress for her weaker students. The CAA system provided a means to give her students the opportunity to practice and to present some of the kind of questions that she wished for her students to be able to answer; however, the system sometimes lacked in this regard, too. Furthermore, like many of the other lecturers, Xi felt that the time investment required to develop new questions made doing so not worthwhile.
Figure H.11: Omicron had two aims for assessment: to encourage students to practise and to give marks to students. The latter aim was imposed by the community rules. She felt that the CAA system questions were not sufficient for her needs.

Nonetheless, towards the close of the interview, Xi remarked about the intention behind the development of the CAA system, which she believed was for the assessment of large groups and providing feedback. She seemed content to continue using the CAA system despite the contradictions she faced.

When asked whether the CAA system is effective, Xi believed it was not. She said, “It’s not. Maybe it could be better, depending on how it’s used. But the way I use it, the way the questions are formulated, the way the students view it, it’s not effective for what I would like in terms of outcomes.”

Xi was the only lecturer to state that the system was not effective, and since she related effectiveness to the fulfilment of outcomes there is little doubt that the assessment was not effective for her aims. However, Xi continued to use the system with her students. This continued to be an unresolved contradiction with no clear indication as to how Xi would act to resolve it. She indicated that she felt she had little option. There remained a need to assess her students and allow them to practise and receive feedback; furthermore, the time saved with these assessments meant that she was able to spend more time on administering other aspects of the course — not least the project assessments that demanded more time.
H.6 Omicron

H.6.1 Subject

Omicron was a mid-career lecturer in the mathematics department whose research expertise was in pure and applied mathematics. While she was primarily a lecturer and researcher in mathematics, she maintained close ties with the mathematics education community at the same institution. Therefore, while she was not primarily concerned with mathematics pedagogy research, she was interested in sharing good practice.

Although it is possible she maintained two identities — as lecturer and as researcher — there was not any evidence of conflicts arising between them in the interviews.

H.6.2 Object

Omicron described her aims for assessment in her own words:

“There are two reasons for giving an assessment. The first reason is to motivate the students to learn the material, and the second reason is that we are required to give a number to each student, indicating how well we feel they’ve done in learning the material...I think most teachers are much more interested in the first thing.”

While she felt an obligation from others to award marks, Omicron made it clear that it did not interest her, describing it as “extremely irritating”. Thus part of the object of her activity was imposed on her by the community of her activity. While this had an impact on the types of assessment she used with her students, she also aimed to provide assessment students that was convenient and motivating.

H.6.3 Tools

For this reason, she maintained her use of CAA. Her purpose for the system was primarily for students to monitor their own progress and they could do so at their own pace and leisure. She also understood the reasons why she believed the students liked the system: as well as the convenience, there is a familiarity with the type of questioning they encountered pre-university — “it’s comforting for them, it’s more like what they do at A-Level”.
Appendix

Omicron was eager to make the assessments as convenient as possible so that students were able to access the tests when they most desired to. She was aware that students were pressured for time:

“They have so many courseworks in their first year. I think they’re over-assessed a bit. Or, at least, they may not be over-assessed but none of the assessments start until week three, and that’s when they all get cramped up: especially certain weeks, they get really quite intense.”

The CAA system was part of a wider assessment strategy for Omicron. She was aware of the shortcomings of the system in terms of assessing conceptual understanding and ensure that students were tested on a breadth of skills: “I want them to be able to carry out the procedures. I want them to be able to do the conceptual things. I just test the procedures through the computer courseworks [CAA tests] and I test the conceptual things through written courseworks. It’s by design, partially, that I’ve done it that way.” She also said, “I’ve intentionally chosen more procedural questions for the computer-aided assessment.”

H.6.4 Community

Omicron described her position in the activity of assessment as though she was a member of the community in the students’ activity of learning. Her practice for assessment was motivated by her consideration of students’ aims.

For example, when asked if she might stop using the CAA system — after having discussed lecturers that had stopped using CAA — the reason she gave for not doing so was related to the students:

“I don’t think so. I don’t think so because I think that the students find it comforting.”

She was concerned that students are in a process of transition as they become accustomed to university learning; offering CAA was one way of facilitating this transition. For her, CAA eased students towards taking responsibility for their own learning and managing their learning time.

Omicron believed that it was important for the students to be able to carry out procedures and to understand the work that they were asked to do. Therefore, she was keen for students to practise and collaborate: “I think for the students to collaborate on things
it’s helpful for everyone. The strong students learn it better if they explain it and the weaker students learn it better if they’ve had it explained more times.”

She was cautious that students might not always collaborate effectively: “If [they] do everything in a collaborative way and they never really check that they themselves have understood it, then they don’t know if they themselves have learned or understood it... It’s quite different to do something in a group and understand what they are saying when they are saying it; but that’s very different from being able to come up with those steps yourself.”

Although Omicron was not clear how students worked on the tests, she had views on how students should work to get best results: “Ideally, in the practice tests, what I would do is have them collaborate for a time and then check some on their own to make sure they really understood it by themselves, before they carry on to coursework [summative test].”

Omicron rarely experienced students asking for support from her, in both tutorial sessions and in the mathematics support centre. She was aware, however, that students sometimes discussed the tests between themselves during tutorials.

She had discussions with other lecturers about CAA practice: some used CAA and others did not. She had talked to Nu about developing questions for the system and to other lecturers (including Mu, who taught another part of the module) about the use of the system more generally. It seemed that these discussions had little impact on her practice, though she suggested there was some consensus that the CAA system should be changed.

Omicron mentioned towards the end of her interview about discussions she had had with a colleague that had been using multiple choice questioning. She suggested that she would be interested in developing more resources for the system, using research to inform her contributions, if she had been given suitable resources:

“I have a friend... who was teaching quite large service courses for a while, and off his assessments were multiple choice. I talked to him about that, and he had some really interesting comments on it: about how to design multiple choice questions in such a way that partial credit was meaningful; that somehow some wrong answers were better than other wrong answers. So, I think there’s quite a bit of interesting research out there, and... if I am one of the people who is put in charge of figuring stuff like this out in the future, and if I’m still teaching [the same module], then I think I’m going to want to go into the literature a little bit and take a look at some of those thing, like what are effective techniques.”
In this quote, Omicron shares a feeling of uncertainty about her status as a teacher of this module, which was in the hands of her department. She had mentioned earlier about a potential restructuring of the mathematics course in the first year. She was concerned that making changes at that stage would have been a waste of time if the course was soon to no longer run:

“If it turns out that [this module] continues and I’m continuing to teach it, then I’ll figure out some way, put in a grant or something...find some way to develop new [questions]. ...At this point I don’t know what’s going on, so I don’t have any plans [to make changes] at the moment.”

**Division of labour**

Omicron acknowledged the work of a colleague who had set up the system some years before and Nu’s work in attempting to develop new questions for the system. However, Omicron felt that the system needed a significant investment in time in order to be able to provide the tests that she wished to give to her students.

She also acknowledged the place of her students in the community of her activity, though she regarded her role as a facilitator in the students’ activity of learning. For her, students should be learning to take responsibility for their own learning and acquiring agency that had been previously held by the teacher.

**Rules**

Omicron communicated the need to abide by the expectations of her department, and in education circles more generally, to provide an outwardly fair mark for each student based on how well they were doing on the course. She did so, though she expressed that it was somewhat of an annoyance when there were other objectives for assessment that she desired to achieve.

Like other lecturers, Omicron was encouraging of collaboration but felt that her students might not always collaborate effectively. She was less concerned that students would plagiarise than other lecturers — her primary focus was to encourage students to practice using and communicating mathematics — and she felt that collaborating was, overall, a benefit to her students.
H.6.5 Contradictions

There were aspects of the system that Omicron mentioned during the interviews that indicated an on-going contradiction with her use of the CAA tests. She expressed a desire to make changes but these were not realistic: “Of course, there are things that I would like to be able to put on there that aren’t there, and to develop a new question under the current system is quite onerous.”

She was keen to make more expansive changes to the system, including an overhaul of the existing bank of questions: “I would love to change the bank of questions and so on; that is quite a labour-intensive process that I frankly don’t have the time for.”

Omicron was also interested in changing the system itself and had made enquiries regarding new CAA systems and the feasibility and affordability of implementing a new system. She believed this desire for change was shared with other lecturers: “I think that other lecturers also feel that a change of system is necessary — at least, from my discussions with them.”

This demonstrates that Omicron was keen on the benefits of CAA, but the contradictions that she faced would be overcome by a change in the CAA system. She was particularly keen on giving students the opportunities to practise, which believed was a particular strength of CAA: “Repetition is the way that you just get that stuff automatic in your mind. And so, I think that this is the best way that we have of presenting to the students several examples — similar examples — that they can just do enough repetition of it that they begin to see how to do it.” While the system in use provided such a facility, Omicron believed that the questions were not sufficient for her means.

H.6.6 Effectiveness

Omicron said there were two reasons for setting an assessment: “The first is to motivate the students to learn the material, and the second reason is that we are required to give a number to each student, indicating how well we feel they’ve done in learning the material...I think that most teachers are much more interested in the first thing, of getting students to understand the material...” She offered this insight into her beliefs for setting assessment as a means to indicate a contradiction. It was an indication that she believed that the CAA system ultimately “gives a number to each student” and does not necessarily “motivate the students to learn the material”, at least in the manner that she intended.

Although Omicron primarily wished to use CAA to help students monitor their own progress, she believed that an assessment would be effective “if it has caused the students
Appendix

Figure H.12: CAA was effective for Omicron’s two main goals for assessment; however she felt the questions were not challenging enough to get a good appreciation of her students’ conceptual knowledge.

to absorb the material that you feel is most important.” In this respect, Omicron believed that the CAA system could be used effectively and that her practice was such a way.

With respect to her aims, the CAA system proved reasonably effective: the practice tests offered the opportunity for practice and repetition; the students could monitor their progress and establish their weak areas to improve upon. She felt that the students liked the CAA tests because “that’s what they’re accustomed to. I think that’s part of the reason it’s comforting for them, it’s more like what they do at A Level.”

While Omicron believed that students were confident with and motivated by the CAA tests, the questions did not post sufficient challenge for her students since their conceptual understanding was not adequately tested:

“The difficulty of the undergraduate curriculum for students is that they’re being asked to move beyond procedures and methods.”

Thus, for Omicron, the students were tested on necessary procedures, but it was the conceptual understanding that was more important to capture in assessment. This on-going contradiction was the motivation for Omicron to look into other CAA systems to establish what facilities they could offer that cannot be achieved with the current system.
References


Davies, P. (2001). Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible? *5th CAA Conference*.


References


