Adaptation of tertiary mathematics instruction to the virtual medium: approaches to assessment practice

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/12561

Publisher: © Sven Trenholm

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

![Creative Commons License](https://creativecommons.org/licenses/by-nc-nd/2.5/)

**Attribution-NonCommercial-NoDerivs 2.5**

You are free:

- to copy, distribute, display, and perform the work

**Under the following conditions:**

**Attribution.** You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or redistribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
Adaptation of Tertiary Mathematics Instruction to the Virtual Medium:
Approaches to Assessment Practice

by

Sven Trenholm

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy of Loughborough University

May 2013

© by Sven Trenholm, 2013
Abstract

Mathematics has been singled out as a challenging discipline to teach fully online (FO). Yet both the demand for and development of FO mathematics courses is increasing with little known about the quality of these courses and many calling for research.

Whereas most research has investigated the nature of these courses by examining instructional outputs such as student grades this research seeks the same insight but by examining instructional inputs. Specifically, it seeks to investigate the nature of current assessment practice in FO mathematics courses.

To conduct this investigation, deep learning (Marton & Säljö, 1976a, 1976b) is used as the principle theoretical framework. From the growing body of literature associated with deep learning, two studies are selected to investigate current FO mathematics instructors’ assessment practices. An additional framework based on empirical findings related to the use of different kinds of feedback is also used. In total, six study measures are used to conduct a mixed methods study in two parts. The target demographic and course context are tertiary instructors from Western nations that teach introductory level mathematics (particularly statistics and calculus). The first study explores current FO mathematics assessment practices using an online survey (n=70) where the majority of participants originate from US higher education institutions. In the second study six of the US survey participants’ are interviewed about how their assessment practices and approaches used in their FO mathematics courses differ from those used in their face-to-face (F2F) mathematics courses.

This study represents the first known attempt to investigate the nature of tertiary FO mathematics instructors’ assessment practices using appropriate theoretical frameworks. In particular, it investigates mathematics instructors’ experiences of the affordances and constraints of the FO course context when adapting their F2F practice to this new environment. Findings suggest the FO course context is a challenging environment for instructors to orient their teaching and assessment practice in a way that helps develop students’ understanding of mathematics. Analysis of interview responses suggests the problem lies with the nature of interactivity provided in the FO course context.
Acknowledgments

I first of all thank God for the opportunity and ability to undertake this study.

It has been a privilege to work with and learn from both of my PhD supervisors, Dr. Carol Robinson and Dr. Lara Alcock. I am very grateful for the many hours of discussion and editing both have provided.

I also wish to thank Alistair, Colin and Dave for shouldering life’s load with me. Our regular times together have been an important anchor point.

I finally wish to acknowledge my wife and six children for their patience, help and support (including many head scratches)! In particular, my wife, Rachel, has patiently endured my ups and downs and provided invaluable encouragement and council to me throughout these years of stress and strain.

Deo sit Gloria

Sven Trenholm
List of Tables

Table 1: Framework presenting S&B study findings ..........................................................56
Table 2: Classifications of mathematics assessment questions compared to Bloom’s taxonomy and S&B’s orientations to assessment practice ..................................................59
Table 3: A proposed taxonomy of kinds of feedback in relation to the effect on learning ....66
Table 4: Summary of Findings from Galante's (2002) Study ...........................................73
Table 5: Summary of Findings from Trenholm's (2007) Study ........................................73
Table 6: Summary of online survey participants’ responses ..........................................89
Table 7: Summary of survey participants’ institutions by country/group ..........................89
Table 8: A framework, based on claims and findings discussed in the literature review, for distinguishing assessment instrumentation according to kind of feedback provided ........94
Table 9: S&B matrix showing scores used to operationalize the findings ........................98
Table 10: Study I research questions and corresponding analysis ................................107
Table 11: Participant demographics ..............................................................................110
Table 12: Assessment instruments used with descriptive statistics ..............................112
Table 13: Use and weighting of discussion for SUNY participants .................................113
Table 14: Use and weighting of quizzes for SUNY participants ....................................114
Table 15: Quality of feedback as measured by the study framework .............................117
Table 16: Timing of feedback .........................................................................................118
Table 17: Use of invigilation with instruments ..............................................................123
Table 18: Participants’ PD Activities ...............................................................................127
Table 19: S&B matrix summarizing participants’ responses with red line viewed as breakpoint separating majority KR from majority KC views ......................................................130
Table 20: Overall summary of participants’ orientations ................................................138
Table 21: Overall scale measure descriptive statistics ....................................................142
Table 22: Correlation matrix of study measures .............................................................146
Table 23: Summary of analysis and findings for Study I - R8 ........................................151
Table 24: Study II research questions and corresponding interview questions and analysis ....171
Table 25: Percentage distribution of US student enrolment in postsecondary education and in DE courses by institution type (2007-08) .................................................................172
Table 26: Demographic make-up of interview candidates ...........................................178
Table 27: Participant approach measures and PD information from the first study (increasing order of S&B measure) ...............................................................................178
Table 28: Summary of interview participants’ use of discussion as a weighted assessment instrument ...................................................................................................................181
Table 29: Summary of participants’ use of quizzes in each context ..............................192
Table 30: Participant use of CAA in F2F vs. FO courses ..............................................193
Table 31: Comparison of F2F and FO course usage of CAA for homework ..................194
Table 32: Instruments that are invigilated by interview participants ............................202
Table 33: Summary of second study results .................................................................239
Table 34: Overall summary of participants’ orientations ...............................................261
Table 35: Orientation of participants’ F2F vs. FO assessment schemes as identified by total assessment weightings (with weighting associated with the use of CAA specified) ............261
List of Figures

Figure 1: The field of e-learning with focus of thesis research in solid boxes ...........................................24
Figure 2: Individual participants’ assessment practices (Note: For this and following figures, where the stacked column is above 100% it is assumed to represent possible bonus points and where the column is below 100% it is assumed to be a participants’ mistake).................................115
Figure 3: Individual participants’ assessment practice by type of feedback used ......................................121
Figure 4: Use of invigilation related to individual participants’ assessment practices (instruments used) ...........................................................................................................................................125
Figure 5: Use of invigilation related to individual participants’ assessment practices (type of feedback used) ...........................................................................................................................................126
Figure 6: Individual participants’ assessment practice related to S&B measure .......................................140
Figure 7: Individual participants’ use of feedback related to S&B measure .............................................141
Figure 8: Scatter plot of CCSF vs. S&B measure with line of regression..................................................144
Figure 9: Scatter plot of ITTF vs. S&B measure with line of regression ..................................................145
Figure 10: Scatter plot of quiz weighting vs. ITTF measure with line of regression .................................149
Figure 11: Screenshot showing Atlas-Ti coded excerpt of interview responses for the first question on use of discussion (Note: A4 and B4 were the original codes used for participants P1 and P2, respectively) .................................................................................................................................164
Figure 12: Screenshot of Atlas-Ti code family manager for the question on the use of discussion. The code families are shown in the upper left-hand corner. (Note: ‘Size’ and ‘Quotes’ refers to the number of codes and quotes, respectively, in a particular family). The list of all codes is shown in the scrollable table in the lower right-hand corner. .........................................................................................................................166
Figure 13: Identified sources of influence affecting the decision to invigilate .......................................203
Figure 14: A representation of Tobin’s (1986) needed timing for the cycle of reflective interactivity for one student as compared to Mandernach, Gonzales, & Garrett’s (2006) expected timing for FO courses (identified need in black vs. FO course in red) .......................................................... 245
Publications

Directly Related to the Thesis


Indirectly Related to the Thesis


## Glossary of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-year Institution</td>
<td>Traditional ‘bricks and mortar’ non-bachelor granting higher education institutions (e.g. North American community college).</td>
</tr>
<tr>
<td>4-year Institution</td>
<td>Traditional ‘bricks and mortar’ bachelor granting (e.g. university, 4 year college).</td>
</tr>
<tr>
<td>Approach Measures</td>
<td>This thesis makes use of three ‘approach measures’. Two come from the Approaches to Teaching Inventory (Trigwell &amp; Prosser, 2004) and one is constructed based on the Samuelowicz &amp; Bain study (2002) framework.</td>
</tr>
<tr>
<td>Approaches to Teaching Inventory (ATI)</td>
<td>A psychometric instrument consisting of two sub-scale measure that are used to measure instructors’ approaches to teaching (Trigwell &amp; Prosser, 2004).</td>
</tr>
<tr>
<td>Blended or Hybrid Courses</td>
<td>These courses use some level of web-mediated instruction to replace traditional instruction (e.g. lecture) but still retain a physical meeting space and time.</td>
</tr>
<tr>
<td>Class vs. Lecture (or Lecture Group)</td>
<td>The terms ‘class’ and ‘lecture’ are considered to be functionally equivalent in the sense that they refer to the group of students formally enrolled in a single course. For F2F courses this group has a regularly scheduled time and place where students meet with the instructor for instruction and assessment. However, for FO courses, this group meets virtually and asynchronously. ‘Class’ is used in the second study because participants tend to prefer this term.</td>
</tr>
<tr>
<td>Computer-Assisted Assessment (CAA)</td>
<td>CAA has been broadly defined as the use of ‘computers to deliver, mark or analyse assignments or exams’ (Sim et al., 2004, p.217). Given the initial computer-mediated state of FO instruction, this study extends the definition to include the use of computer automation in question generation and/or grading.</td>
</tr>
<tr>
<td><strong>Computer-Assisted Assessment (CAA) Feedback</strong></td>
<td>Feedback that is generated and provided, using some combination of computer software and hardware, in response to submitted answer(s) for an assessment question or instrument. It is generally automated but may be set by the instructor (e.g. kind of feedback provided, timing of the feedback). Also referred to as feedback via <em>computer agency</em> or <em>computer-generated feedback</em>.</td>
</tr>
<tr>
<td><strong>Conceptual Change/Student-focused (CCSF)</strong></td>
<td>One of the two ATI sub-scale measures. It provides a measure of how instructors’ approaches are oriented to a ‘student-focused strategy aimed at students changing their conceptions’ (Trigwell &amp; Prosser, 2004, p. 413).</td>
</tr>
<tr>
<td><strong>Conference Board of Mathematical Sciences (CBMS) Survey</strong></td>
<td>Every five years since 1990 the Conference Board of Mathematical Sciences (CBMS) has sponsored a national survey of undergraduate mathematical and statistical sciences in US four-year and two-year universities and colleges. This survey is published by the American Mathematical Society. CBMS is an umbrella organization consisting of sixteen professional societies all of which have as one of their primary objectives the increase or diffusion of knowledge in one or more of the mathematical sciences (<a href="http://www.cbmsweb.org/">http://www.cbmsweb.org/</a>).</td>
</tr>
<tr>
<td><strong>Course vs. Module</strong></td>
<td>The term ‘course’ will be used throughout this paper. The equivalent UK term is ‘module’. ‘Course’ is used in the thesis because most participants are from the US.</td>
</tr>
<tr>
<td><strong>Distance Education (DE)</strong></td>
<td>DE has been broadly characterized as education that involves the physical separation of instructors and learners. Where this education includes separation in space <em>and</em> time it is referred to as ‘asynchronous’. Where it simply involves separation in space it is referred to as ‘synchronous’ (Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004). Currently the most prevalent form of DE is asynchronous (e.g. Sumler, 2001).</td>
</tr>
<tr>
<td><strong>Face-to-Face (F2F)</strong></td>
<td>Refers to the form of instruction that is traditionally experienced at the tertiary level (e.g. lecture). That is, instruction takes place whereby students and the instructor(s) are scheduled to physically meet together at a regular time and place.</td>
</tr>
<tr>
<td>Feedback Measures</td>
<td>Three measures created based on findings in the literature and used to analyze instructors’ feedback practices. The measures refer to three different kinds of feedback used by instructors: ‘correct/incorrect’, ‘full solution’ and ‘hints and comments’.</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Fully Online (FO) Instruction</td>
<td>Refers to fully asynchronous online instruction. That is, instruction whereby students and the instructor(s) do not physically meet together at the same time and same place.</td>
</tr>
<tr>
<td>Information Transmission/Teacher-focused (ITTF)</td>
<td>One of the two ATI subscale measures. It provides a measure of how instructors’ approaches are oriented to ‘teacher-focused strategy with the intention of transmitting information to students’ (Trigwell &amp; Prosser, 2004, p. 413).</td>
</tr>
<tr>
<td>Instructor vs. Teacher</td>
<td>The term ‘instructor’ will be used in reference to the person or people involved in providing the course instruction. According to the terminology used in the study relevant quotes may refer to ‘lecturer’, ‘teacher’, ‘professor’, academic’, ‘staff’ … In the context of this study, these roles are considered to be functionally equivalent to the role of an instructor.</td>
</tr>
<tr>
<td>Invigilation vs. Proctoring</td>
<td>These are equivalent terms referring to human supervision of student assessment. ‘Invigilation’ is commonly used in the UK instructional context whereas ‘proctoring’ is commonly used in the US instructional context.</td>
</tr>
<tr>
<td>Knowledge Reproduction (KR) to Knowledge Construction/Transformation (KC)</td>
<td>The two end-points of Samuelowicz and Bain’s (2002) continuum representing the variation in instructors’ orientations to assessment practice.</td>
</tr>
<tr>
<td>Open University</td>
<td>An institution of higher education that specializes in providing distance higher education.</td>
</tr>
<tr>
<td>Samuelowicz and Bain (S&amp;B)</td>
<td>Samuelowicz and Bain’s (2002) study that identified instructors’ ‘orientations to assessment practice’.</td>
</tr>
<tr>
<td>State University of New York (SUNY)</td>
<td>The State University of New York (SUNY) is the umbrella term for the state of New York’s system of public institutions of higher education. It comprises of 29 community colleges (i.e. 2-year institutions) and 35 universities (i.e. 4-yr institutions).</td>
</tr>
<tr>
<td>Study Measures</td>
<td>Constitute the three feedback measures and the three approach measures.</td>
</tr>
<tr>
<td>Virtual Learning Environment (VLE)</td>
<td>The web-based environment used to host fully online and blended or hybrid courses.</td>
</tr>
</tbody>
</table>
# Table of Contents

Abstract .................................................................................................................................................. 2
Acknowledgments ...................................................................................................................................... 3
List of Tables ........................................................................................................................................... 4
Publications .............................................................................................................................................. 6
Glossary of Terms ..................................................................................................................................... 7

Table of Contents ................................................................................................................................. 11

1. Introduction ......................................................................................................................................... 17
   1.1 Background of the Study ................................................................................................................ 17
   1.2 Purpose of the Study ....................................................................................................................... 17
   1.3 Significance of the Study ................................................................................................................ 18
   1.4 Theoretical Framework .................................................................................................................. 18
   1.5 Research Questions ....................................................................................................................... 19
   1.6 Delimitations .................................................................................................................................. 21
   1.7 Overview of Thesis ....................................................................................................................... 21

2. Overview of FO Instruction within E-Learning .................................................................................. 23
   2.1 Background: Definitional Issues and Growth ............................................................................. 23
   2.2 General Potential Benefits ........................................................................................................... 25
   2.3 Current State of Research ............................................................................................................ 26
   2.4 Meta-Analyses Comparing General F2F to FO Instruction ....................................................... 27
      2.4.1 Research Quality ..................................................................................................................... 28
   2.5 Summary ....................................................................................................................................... 31

3. The Nature of Mathematical Knowledge, Instruction and Assessment ............................................. 32
   3.1 Previous Meta-Analysis from a Mathematics Perspective .......................................................... 32
   3.2 Nature of Mathematical Knowledge ............................................................................................. 33
   3.3 Instruction and Assessment in Mathematics Courses ................................................................... 34
      3.3.1 Instruction in Mathematics Courses ....................................................................................... 35
      3.3.2 Assessment Practices in Mathematics Courses ..................................................................... 36
   3.4 Summary ....................................................................................................................................... 37

4. Fully Online Mathematics ................................................................................................................. 38
   4.1 Web-Assisted vs. Web-Based Distinction ...................................................................................... 38
   4.2 Overview of Current FO Mathematics Research ......................................................................... 39
4.3 Attrition in FO Mathematics Courses .................................................................40
4.4 Five Issues of ‘Fit’ .............................................................................................41
  4.4.1 Constructivist Pedagogy .............................................................................42
  4.4.2 Student-Led Learning ..............................................................................42
  4.4.3 Use of Discussion .....................................................................................43
  4.4.4 Use of E-Lectures ....................................................................................45
  4.4.5 Communicating Mathematics ...................................................................46
4.5 The Quality of Learning in the FO Instructional Context ..............................47
4.6 Summary ..........................................................................................................49

5. Research on Approaches to Teaching and Assessment ....................................50
  5.1 Broader Theoretical Background: Deep Learning .........................................50
    5.1.1 History .....................................................................................................50
    5.1.2 The Quality of Learning in the Mathematics Instructional Context ..........51
  5.2 Instructors’ Approaches ................................................................................52
    5.2.1 Importance of Investigating Instructors’ Approaches .........................52
    5.2.2 Approaches to Teaching in General ......................................................53
    5.2.3 Approaches to Teaching Mathematics ................................................54
    5.2.4 Approaches to Assessment in General ..............................................54
    5.2.5 The ATI compared to the Samuelowicz and Bain 2002 Study ............57
    5.2.6 Approaches to Assessment in Relation to Mathematics Questions .......58
  5.3 Summary ..........................................................................................................60

6. The Practice of Assessment in FO Mathematics Courses ...............................61
  6.1 Theoretical Background: Assessment Practice ............................................62
  6.2 The Practice of Feedback in Assessment ......................................................63
    6.2.1 Research Overview ..............................................................................63
    6.2.2 Feedback Timing ..................................................................................64
    6.2.3 Feedback Kind .....................................................................................64
  6.3 F2F Mathematics Instructors’ Assessment Practices – Instruments, Weighting and Delivery .................................................................66
  6.4 FO General and Mathematics Instructors’ Assessment Practices ....................69
    6.4.1 General FO Instructors’ Assessment Practices ....................................69
    6.4.2 General FO Instructors’ Assessment Feedback Practices ....................70
    6.4.3 FO Mathematics Instructors’ Assessment Practices .............................71
    6.4.4 FO Mathematics Instructors’ Assessment Practices Identified for Investigation...73
12.1 How and why is discussion/interaction used? ................................................................. 242
   12.1.1 Limitations ............................................................................................................. 246
   12.1.2 Summary ............................................................................................................... 247
12.2 How and why are quizzes being used? .............................................................................. 247
   12.2.1 Limitations ............................................................................................................. 250
   12.2.2 Summary ............................................................................................................... 251
12.3 How and why are participants choosing to use invigilation? ........................................... 251
   12.3.1 Summary ............................................................................................................... 253
12.4 How and why is feedback being provided? ....................................................................... 253
   12.4.1 Limitations ............................................................................................................. 259
   12.4.2 Summary ............................................................................................................... 260
12.5 How are participants’ approaching their FO course assessment practice? ...................... 260
   12.5.1 Limitations ............................................................................................................. 263
   12.5.2 Summary ............................................................................................................... 263
12.6 Summary of Research Findings ....................................................................................... 264
13. Conclusions .......................................................................................................................... 267
   13.1 Implications for Practice .............................................................................................. 269
   13.2 Recommendations for Further Research ..................................................................... 270
   13.3 Concluding Statement ................................................................................................. 272
References .................................................................................................................................. 273
Appendices ............................................................................................................................... 310
1. Introduction

1.1 Background of the Study

Faced with an increased demand for higher education, institutions and government policy makers are turning to fully online (hereafter termed FO) courses as a viable, effective and efficient means of delivering instruction (Miller, 2010, Bakia, Shear, Toyama, & Lasseter, 2012). As the name suggests, formal instruction in these courses is completely mediated by the virtual Internet medium. Growth in these courses is considered to be ‘exploding’ (Campbell, 2012). For example, in the US, ‘online’ enrolments at degree-granting tertiary institutions grew at an average annual rate of about 14% for the five years from 2006 to 2010 (Allen & Seaman, 2011) with continued growth projected (Christensen, Horn, Caldera, & Soares, 2011)

Accordingly, both the demand for (Selden, 2005) and development of FO mathematics courses is increasing (Maltempi & Malheiros, 2010; Zinger, 2006) and substantial investments of both time and money are being made in developing these courses (Engelbrecht & Harding, 2005b). Currently the US Conference Board of Mathematical Sciences (CBMS; 2010) survey reports that 35% of four-year mathematics departments offered distance-learning courses of which 72% is ‘completely online’ and 88% of two-year colleges offered distance-learning courses of which 73% is ‘completely online’ (Kirkman, 2012, Table SP.10).

Within this climate mathematics has been singled out as a challenging discipline to teach in this modality (Engelbrecht & Harding, 2005b; Glass & Sue, 2008; Lokken, 2011; Philip, 2003; Smith, Ferguson, & Caris, 2003; Smith & Ferguson, 2005; Zinger, 2006) and little attention is being paid to disciplinary characteristics in current FO course development (Smith, Torres-Ayala, & Heindel, 2008).

1.2 Purpose of the Study

Some research has investigated the nature of learning in these courses as they compare to traditional lecture courses (what will be termed face-to-face or F2F). Almost all this research involves some form of comparison in student achievement measures (e.g. students’ final exam grades in FO vs. traditional lecture courses; e.g. Xu & Jaggars, 2011). This research also seeks to investigate the nature of these courses. However, rather than focusing on what may be termed the output elements of FO mathematics courses, such as student grades, this research focuses on input elements related to instructor practices. In particular, given claims that
assessment is a powerful influence in directing student learning (Davis et al., 2005; Houston, 2002; Marriott & Lau, 2008; Smith & Wood, 2000), this study seeks to investigate the nature of these courses through the lens of FO mathematics instructors’ assessment practices.

1.3 Significance of the Study

In view of current and projected growth of FO courses offered at tertiary degree-granting institutions and considering current challenges in teaching FO mathematics courses as well as the fact that little is known about the nature of these courses, many are calling for more research (Engelbrecht & Harding, 2005b; Maltempi & Malheiros, 2010; Mills & Raju, 2011; Montiel & Bhatti, 2010).

The little research that has investigated both tertiary F2F and FO mathematics instructors’ assessment practices has largely focused on individual assessment instruments and course assessment schemes. This study represents the first known attempt to investigate the nature of tertiary mathematics instructors’ assessment practices using appropriate analytical frameworks. In particular, this study investigates these instructors’ experiences of the affordances and constraints of the FO course context when adapting their F2F practice to this new environment.

Findings from this study will help elucidate some of the reasons why teaching mathematics at a distance and FO has historically proved to be challenging. They will also provide, for example, course developers and policy makers information that will help address the discipline-specific needs of mathematics.

1.4 Theoretical Framework

A body of research on approaches to teaching (e.g. Even & Kvatsinsky, 2009), and more recently approaches to assessment (e.g. Samuelowicz & Bain, 2002), has recently emerged in the education literature. These studies have evolved out of Marton and Säljö’s (1976a, 1976b) early work on ‘deep learning’, which describes qualitative differences in learning outcomes. In simplest terms, students’ approaches to their learning were found to be somewhere on a continuum from surface to deep approaches. Studies on instructors’ approaches similarly seek qualitative measures of how instructors view, and for what purpose they use assessment instruments in their courses. This research on instructors’ approaches presents viable methods of investigating the nature of FO instructors’ teaching and assessment practices.
From this growing body of research two studies were initially selected as a basis for analyzing FO mathematics instructors’ approaches to teaching and assessment:


The former study presents an established psychometric instrument used to measure the ‘variation between an information transmission/teacher-focused view of teaching and a conceptual change/student-focused view of teaching’ (Trigwell & Prosser, 2004, p.415). The latter study, hereafter termed the S&B study, provides an initial framework for analysis which is then also used to create a summated scale measure of instructors’ approaches to assessment.

A third analytic method is also introduced. This framework is created based on a review of the literature on feedback practices, where the kind of feedback (e.g. right/wrong, full solution, hints) instructors provided was found to have a significant effect on student learning. In the end, this study investigates instructors’ approaches to teaching and assessment practice using one established psychometric instrument (ATI), a framework and novel summated scale based on the S&B study findings and a further novel framework created to analyze instructors’ feedback practices.

1.5 Research Questions

The main overarching research question asks: what is the nature of current FO mathematics courses? This question is addressed by a mixed methods study conducted in two parts. The first study asks what specific assessment practices are used in these courses and whether some of these practices are in any way related to measures of instructors’ approaches to teaching and assessment. The first study considers eight separate questions.

*Study I*

R1. What instruments are FO mathematics instructors currently using to assess their students? How are these weighted?

R2. How are instructors using feedback in their FO mathematics courses?
R3. How are instructors using invigilation in their FO mathematics courses?

R4. What kind of professional development opportunities are FO mathematics instructors receiving for their courses?

R5. Using the findings of Samuelowicz and Bain’s (2002) ‘Identifying academics’ orientations to assessment practice’ study as a framework, how are FO mathematics instructors approaching assessment in their courses?

R6. How are FO mathematics instructors approaching teaching in their courses as measured by Prosser and Trigwell’s (2004) Approaches to Teaching Inventory (ATI)?

R7. How do findings in question six relate to findings in question five?

R8. Are there any statistically significant differences in any of the study measures based on usage of invigilation, a greater variety of instruments, quizzes or discussion (the latter two as weighted instruments)? When used, is the weighting of either quizzes or discussion related to any of the study measures?

These questions are answered using responses from an online questionnaire completed by 70 FO mathematics instructors who mostly teach in US institutions. Participants’ assessment practices are detailed. Measures of participants’ approaches to teaching, assessment and feedback are determined and statistical analysis is used to explore whether these measures are associated with specific assessment practices identified in the literature.

Directed by the first study, the second study asks how and why some assessment practices are used, how individual participants are approaching their assessment practice and how do these results reflect on the quality of learning in FO courses? Five separate research questions are posed.

**Study II:**

R1. How and why is discussion/interaction used?

R2. How and why are quizzes being used?

R3. How and why are participants choosing to use invigilation?

R4. How and why is feedback being provided?
R5. How are participants’ approaching their FO course assessment practice?

These questions are answered via semi-structured interviews with six of US survey participants who took part in the first study and all teach in US public higher education. Participants are asked to compare assessment practices and approaches used in their FO mathematics courses to those used in their F2F mathematics courses and this data is analyzed using the constant comparative method (Boeije, 2002) to build explanations regarding the nature of current FO assessment practice.

From a US public higher education perspective, the thesis research provides empirical findings that help explain why teaching mathematics in the current FO course context has proved to be so challenging. Findings from this study question whether FO mathematics courses can be taught and assessed for depth of understanding without significant changes in current FO course development.

1.6 Delimitations

The scope of the thesis is limited in two ways. First, the target course context for this research is introductory level mathematics (introductory statistics and calculus in particular) where most FO mathematics courses are currently offered (Kirkman, 2012). Second, given this researcher’s familiarity with the US FO mathematics instructional community, the target demographic are tertiary instructors from the US and other Western nations. In particular, the second study focuses on instructors who teach introductory level courses in the US public higher education context.

1.7 Overview of Thesis

The thesis begins with the literature review consisting of five chapters. First, claims and empirical findings in the general FO instruction literature are summarized and research needs relevant to the thesis are identified. Second, one of these needs is addressed with a review of the literature related to the disciplinary characteristics of mathematics and F2F mathematics instruction. Third, the prior review is contrasted with the literature on effective FO instruction and it is argued that the current disciplinary culture of mathematics is not aligned with the current culture of FO instruction. Fourth, the literature on teaching and assessment approaches is reviewed and two studies are identified for use in the thesis: Trigwell & Prosser’s (2004) ‘Development and use of the approaches to teaching inventory’ study and Samuelowicz &
Bain’s (2002) ‘Identifying academics' orientations to assessment practice’ study. Finally, FO mathematics assessment practice is identified as a substantive area of research and the literature with respect to tertiary mathematics assessment practice (both F2F and FO) is summarized and a third method of analysis is presented.

Following the literature review the next six chapters present the two thesis studies. An introductory chapter first describes the mixed methods approach used in the thesis. The next chapter is devoted to detailing the first study methods and methodology. This is followed by the second chapter detailing the study results together with discussion. This chapter is subdivided into three parts – demographic information, assessment specifics and teaching and assessment approaches. The second study, where a sample of US survey participants are interviewed, is presented in the next three chapters. In one chapter the methods and methodology are detailed. The results and discussion follow in two separate chapters.

The thesis ends with a conclusion and implication for practice and future research. Finally appendices and references are provided.
2. Overview of FO Instruction within E-Learning

The following section provides a brief overview of some of the research on ‘e-learning’ (short for ‘electronic-learning’) followed by a review of current research on general FO instruction. To begin, FO instruction is defined within the broader field of e-learning. This is followed by a summary of potential benefits of FO instruction. Background literature on FO instruction is provided by means of a review of two meta-analytic studies comparing traditional F2F and FO instruction. Finally, the current research is critiqued and a summary list of problems – both general and those specific to this study – is discussed. While some of this research comes from secondary instruction, almost all the research cited in this thesis focuses on the tertiary level. From these specific issues, a preliminary rationale for the thesis is provided.

2.1 Background: Definitional Issues and Growth

The term e-learning is still not yet clearly defined. For example, while e-learning has been simply defined as the use of any computer technology to facilitate learning (Shih, Feng, & Tsai, 2008), it has also more precisely been defined as ‘content and instructional methods delivered on a computer (whether on CD-ROM, the Internet, or an intranet), and designed to build knowledge and skills related to individual or organizational goals’ (Clark, 2002, p.2). These definitions, which are fairly wide in scope, may be summarized as learning that results from instruction that makes use of computer software, hardware and/or the internet, to support or partially or completely replace F2F instruction. Within e-learning, FO instruction is viewed as a complete or potentially complete replacement for traditional F2F instruction (see Figure 1). Allen and Seaman (2008), for example, in their major US survey, define ‘online’ courses as those with more than 80% of the content delivered online. Such imprecise definitions – with somewhat arbitrary percentages – are reflective of some of the challenges in defining research in this area.

With regards to the focus context of this thesis, the FO instructional context has two defining characteristics: First, it does not require the instructor(s) or students to meet in the same physical location. Second, this instruction is considered to be primarily ‘asynchronous’ – where ‘synchronous’ instruction may be defined as ‘simultaneous or ‘real-time’ computer-based instruction’, ‘asynchronous’ instruction is ‘not based on simultaneous computer-based instruction’ (Parsad, Lewis, & Tice, 2006, p.14). Considered together, these characteristics imply that instructor(s) and students are not required to meet at the same place or the same time.
Regarding the thesis research the term ‘FO’ refers to this form of instruction. However, regarding the literature review, the term ‘FO’ is also used although it is not always clear the research being reviewed focuses on instruction that is asynchronous. For example, there often appears to be a tacit assumption of asynchronicity (a popular slogan for this form of instruction is ‘anytime and anywhere’; cf. Allen & Seaman, 2008).

Figure 1: The field of e-learning with focus of thesis research in solid boxes

Growth in this kind of instruction is very strong. Allen and Seaman’s (2011) recent US survey, for example, investigated ‘online’ enrolments at degree-granting tertiary institutions using responses from more than 2,500 colleges and universities. For the five years from 2006 to 2010, they found an average annual growth rate of about 14%. Similarly, from 2005 to 2010, the US Conference Board of Mathematical Sciences (CBMS) 2010 survey, using stratified random sampling with sub-population response rates above 54%, investigated enrolment growth in tertiary mathematics distance learning courses. Here they define a ‘distance learning course [as one where] the instruction occurs with the instructor and the students separated by time and/or
place (e.g. where the majority of the course is taught online, or by computer software, by
television or by correspondence)’ (Kirkman, 2012). The recent 2010 study found that 35% of
four-year mathematics departments offer distance-learning of which 72% is ‘completely online’
and 88% of two-year colleges offer distance-learning of which 73% is ‘completely online’
(Kirkman, 2012, Table SP.10). Moreover in the previous survey it was found, for example, that
from 2000 to 2005 university distance education course enrolments in Calculus I and elementary
statistics grew by approximately 300% (Kirkman, Lutzer, Maxwell, & Rodi, 2007).

The focused field of this study and the field of ‘distance education’ (DE) increasingly share a
common body of research literature. At the same time the notion of which students take these
courses appears to be changing. That is, although the CBMS report uses the term ‘distance
education’, the findings are relevant because technological advancements mean the field of DE
is now dominated by FO instruction (Bernard, Abrami, Lou, & Borokhovski, 2004). In addition,
the situation is such that the traditional notion of distance in DE has become obscured and some
see a merging of ‘remote' DE and ‘local’ F2F student populations (Woo et al., 2008). The
result, and what may explain at least some of the growth in this modality, is that increasing
numbers of on-campus or traditional F2F students are also taking FO courses.

2.2 General Potential Benefits

Framing any investigation into potential benefits (and problems) is the relative youth of this
modality coupled with the ongoing advances in technological capabilities (i.e. both at the hard
and software-ends, with improvements in broadband access and speed fuelling greater
development of more bandwidth-hungry and complex systems; e.g. Faulin et al., 2009). In this
respect, the potential benefits of FO instruction listed below represent only a snap shot in time.
They include:

1. Improved instruction through efficient use of feedback mechanisms (Swan, 2003).
2. Improved instruction through the use of adaptive systems that individualize the
   instructional approach (Means, Toyama, Murphy, Bakia, & Jones, 2009; Swan, 2003).
3. Increased access to instruction for disadvantaged and rural groups of students (Bell,
   2010).
4. Increased access to an ‘array of high-quality, interactive learning materials and activities’
   (Mayadas, Bourne, & Bacsich, 2009; Swan, 2003).
5. Some suggestion, when compared to F2F learning outcomes, of being beneficial to deeper conceptual (vs. shallower procedural) understanding (Parker & Gemino, 2001).


From this list the ideas of efficiency, access and depth of learning emerge. In support of some of these claims, the US National Center for Academic Transformation (NCAT), in their action research-based course redesign projects (from 1999 to 2004), found that 25 of 30 course redesign projects showed ‘significant increases in student learning... (with) all thirty institutions reduc(ing) their costs by 37% on average’ (http://www.thencat.org/). However, while efficiency and access issues seem more apparent, the quality or depth of learning is less clear. Additionally, little is known regarding how these benefits may differ among different academic disciplines.

2.3 Current State of Research

In the current FO research context, many in the research community are calling for more research to investigate FO instruction (Bernard, Abrami, Lou, & Borokhovski, 2004; Li & Irby, 2008; Lockee, Moore, & Burton, 2001; Merisotis & Phipps, 1999; Rovai, 2003; Swan, 2003; Tallent-Runnels et al., 2006). However, despite these calls, the growth of online instruction shows ‘no signs of slowing’ and some appear to operate with the thinking that research backing can wait (Allen & Seaman, 2008, p.1). This growth may be aided, for example, in the US, by the oft-cited and critiqued (e.g. Lockee et al., 2001) ‘No Significant Difference Phenomenon’ website (http://www.nosignificantdifference.org/) – that has compiled 355 research reports that support the title’s claim of no significant differences in F2F vs. FO course learning outcomes. Others have argued that a prevailing climate of ‘technopositivism’ (Njenga & Fourie, 2010) continues to be influential. For example, multimedia benefits are discussed as ‘intuitively correct’ (Clark & Feldon, 2005, p.3) or as ‘intuitively appealing... [and] a triumph of enthusiasm over substantive examination of structural processes in learning and instruction’ (Clark, 1994, p.5). It may equally be true that the potential benefits present a rationale for advancing FO instruction. In balance, the jury is still out on whether what we are experiencing is what can be described as ‘transformative’, ‘a paradigm change’ or ‘revolutionary’ (Bourne, Mayadas, & Moore, 2004), and there is a clear need to investigate FO pedagogy.
2.4 Meta-Analyses Comparing General F2F to FO Instruction

To help frame the research on FO instruction, this section summarizes the findings from two recent meta-analyses. Both meta-analyses attempted to measure the effectiveness of FO instruction by investigating studies that compare F2F to FO instruction.

In the first meta-analysis Bernard, Abrami, Lou, Borokhovski, Wade et al. (2004) investigated asynchronous and synchronous DE compared to F2F instruction. They examined over 650 empirical studies of DE constituting a range of technological developments from the use of hypertext systems to present day web-based interactive multimedia. Of these, 167 studies were selected that met ten strict inclusion/exclusion criteria. The study looked at the effect sizes of DE using a statistical formula to provide a measure of comparison (of achievement, attitude and retention) in DE (as the experimental group) vs. F2F classrooms (as the control group). In terms of achievement, the study found a ‘very small and significant effect favouring…DE (overall)’ with asynchronous DE instruction being favoured over F2F while F2F is favoured over synchronous DE. Taking into account the generally poor overall quality of study designs, they concluded that effect sizes were ‘essentially zero’ (p.379). The most significant finding was the extremely wide effect size variability, suggesting that while some DE courses work very well, others perform extremely poorly.

In the second meta-analysis, Zhao et al. (2005) investigated FO instruction compared to F2F with the purpose of identifying factors that influence the effectiveness of FO instruction (contrasted with the previous meta-analysis which was concerned with measuring the effectiveness of DE). They identified 8,840 potential articles of which only 51 were selected on the basis that they were deemed to contain sufficient information to effectively calculate the effect size (for a total of 98 effect sizes from 51 studies and 11,477 participants). Their effect size was ‘a measure of standardized mean difference … computed to estimate the extent of the difference between online learning and face-to-face learning’ (p.1848). They also found that there was no overall significant difference with F2F, and concluded that several factors appear to lead to better outcomes. Among these were the importance of interaction and ‘live’ instructor involvement, as well as the suggestion that FO instruction ‘may be more effective in teaching some content than others’ (Zhao et al., 2005, p.1864). As will be discussed later, the latter factor is considered relevant given the disciplinary emphasis of the present study.
While, on the surface, both these studies support the idea that the effect of FO instruction is not different from F2F, the truth is much less clear. Both studies affirm the following:

1. As reflected in the article selection process, much of the research is substandard (Bernard, Abrami, Lou, & Borokhovski, 2004; Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004; Merisotis & Phipps, 1999); a situation that Bernard, Abrami, Lou, & Borokhovski (2004) have referred to as a ‘methodological morass’ (p.1).

2. Of those studies deemed acceptable for inclusion, there was what Bernard, Abrami, Lou, Borokhovski et al. (2004) describe as ‘wide and unexplained variability’ in outcomes (p.406) and what Zhao et al. (2005) describe as ‘remarkable difference(s)’ in outcomes (p.1861).

It has been suggested that this variation is comparable to what may be found in F2F instruction (Mayadas et al., 2009). However, these meta-analyses appear to refer to variation beyond what may be expected in a F2F setting. This argument is supported by a recent Science magazine review of online education (which also includes partial F2F or ‘blended’ courses). In it, while lauding the benefits of online education, the authors are careful to use the word ‘equivalent’ outcomes for ‘well-designed online courses taught by experienced instructors’ (italics mine; Mayadas et al., 2009, p.86).

### 2.4.1 Research Quality

As the previous meta-analyses point out, there are significant and persistent methodological issues in much of the current research. The following is a summary of shortcomings:

1. Failure to build ‘appropriately on existing knowledge and theories’ (Conole, 2004, p.3).

2. Confusion regarding what is being studied (i.e. the method vs. the medium\(^1\); Merisotis & Phipps, 1999; Rovai, 2003; Tallent-Runnels et al., 2006) and over-simplistic research approaches that bluntly look at ‘the delivery medium’ as the treatment variable and ‘student achievement, or learning’ as the dependent variable (Lockee et al., 2001, p.60).

---

\(^1\) Recognized as a serious issue in much of the research on educational media, the ‘media vs. method’ argument contends that learning is ultimately only affected by the method of instruction not the media used to instruct. The best known proponent of this argument is Dr. Richard E. Clark. He argued that researchers, confusing the medium (or media) with the method (or pedagogy), fail to recognize that “the instructional method is the ‘active ingredient’, not the medium—the medium is simply a neutral carrier of content and of method” (Clark, 1983,1984 as cited in Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004, p.381).
3. A tendency to ‘rely on the use of limited self-reports and qualitative’ vs. more objective measures and quantitative evaluations (Rovai, 2003, p.111).

4. Poor research design, involving:
   a. Failure to control for extraneous variables; e.g. student demographic characteristics (e.g. academically prepared vs. underprepared; Jaggars & Bailey, 2010);
   b. Lack of random selection;
   c. Poor or no reliability and validity for the instruments;
   d. Failure to control for attitudes and beliefs of students and faculty causing reactive effects (e.g. novelty or John Henry effect; Lockee et al., 2001; Mayes, 2004; Merisotis & Phipps, 1999).

5. Use of quantitative achievement measures such as, for example, grades on single tests or grades derived from a combinations of instruments that do not necessarily reflect the quality of learning outcomes (Morrison, Ross, Gopalakrishnan, & Casey, 1995).

6. The disciplinary context is largely ignored (Anderson & Elloumi, 2008; Bates & Poole, 2003; Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004; Conole, 2004; Fardon, 2003; Northedge & McArthur, 2008).

The last two issues, perhaps the most salient with regards to the present research focus, are now discussed further.

**Quantitative vs. Qualitative Outcome Measures**

Learning effectiveness is considered as a measure of whether FO learning is equivalent to or better than F2F learning (Swan, 2003). It is measured in the first meta-analysis using objective measures such as ‘standardized tests, researcher-made or teacher-made tests, or a combination of these’ (Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004, p.390). The second meta-analysis used ‘indicators of effectiveness’ such as ‘one or more of the following measures: grades, quizzes, independent/standardized tests, student satisfaction, instructor satisfaction, dropout rate, student evaluation of learning, student evaluation of course, and external evaluation (where) grades usually are the final scores students received for the class’ (Zhao et al., 2005, p.1844). In either case, increases in quantitative measures of ‘achievement’ or ‘performance’ may not be directly correlated with the quality of learning. In mathematics, for example, higher levels of achievement may reflect only better surface level understanding (Hernandez-Martinez et al., 2011).
Furthermore, achievement may be differentiated based on the quality of learning being addressed by individual questions. For example, Parker and Gemino (2001) compared performance overall and by question of final exams taken by F2F (n=107) vs. FO (n=128) students in a third year course in Systems Analysis and Design. While they found no significant difference in final exam scores, they also found F2F students scored significantly higher on the application of technique part of the exam, while FO students scored significantly higher on the conceptual part of it.

In the end, unless the type of ‘achievement’ is clearly defined and the instrument(s) used to operationalize it are shown to be valid and reliable, it is difficult or impossible to interpret, for example, exactly what ‘effective learning’ might mean. As Micari, Light, Calkins and Streitwieser (2007) sought in their study: ‘it is not merely performance but also how learners think and how their thinking changes that we should be measuring’ (p.458).

**The Lack of Disciplinary Coverage**

Another important issue is that most of the research on FO instruction comes from a de-disciplined research perspective. Where meta-analytic findings fail to make clear distinctions along disciplinary lines, findings may be (erroneously) extrapolated across all disciplines by, for example, instructional designers or policy makers. Indeed, the disciplinary context receives little attention in current DE research (Zawacki-Richter, Baecker, & Vogt, 2009) – a matter that has particularly been raised regarding the FO mathematics research literature (Smith, Torres-Ayala et al., 2008).

This lack of discipline-specific research has been noted in the general F2F context (Lindblom-Ylänne, Trigwell, Nevgi, & Ashwin, 2006) where disciplinary differences are seen to be largely overlooked in formulating policy and practice (Becher, 1994; Neumann, 2001). The importance of the disciplinary context has been highlighted in recent studies. Lattuca and Stark’s (1995) study, for example, showed that the epistemology of each discipline has a strong effect on teaching practice and is thus an important consideration in pedagogical development. Additionally, Lueddeke’s (2003) findings from his exploratory study of disciplinary differences suggest that the disciplinary context has the ‘strongest influence’ on how teaching is conceptualized while ‘qualifications and years of teaching have a moderate impact, and gender

---

2 Different terminology, used in the research literature, for the same construct include ‘enhanced’ (e.g. Butler & Roediger III, 2007), ‘efficient’ (e.g. Stricker, Weibel, & Wissmath, 2011) or ‘effective’ learning (e.g. Karr, Weck, Sunal, & Cook, 2003).
and post do not appear to play a significant part’ (p.213). Instead, current research and development has sought to ‘impose uniformity of approach... [that is seen to] constrain the necessary diversity involved in fitting teaching or assessment methods to... contrasting subject areas’ (Entwistle, 2009, p.150). Or as results from Jones’ (2009) recent study, contrasting five disciplinary areas (outside of mathematics), suggests: ‘a de-disciplined approach...has led to problems in the areas of educational policy and practice’ (p.85). In short, attention to the disciplinary context is considered vital to any educational or pedagogical research or development.

This lack of discipline-specific research has also been noted with regards to educational innovations. Neumann, Parry and Becher (2002), for example, conclude that there is ‘a similarly indiscriminating eagerness to embed methods found effective in one disciplinary area... in other less amenable subject fields before the consequent disappointment and the related acknowledgement of unsuitability set in’ (p.414). In summary, disciplinary differences have generally been ignored (Conole, 2004) and as Becher (1994) writes, there is a tendency to impose uniform approaches even when it is clear the approach is inappropriate.

The next two chapters will address this issue. Chapter three reviews the literature on current disciplinary differences and characteristics related to mathematics. From this basis, chapter four contrasts these findings with the current characteristics of FO instruction where, as the balance of evidence strongly suggests, the extrapolation of general findings to mathematics is not justified.

2.5 Summary

- FO instruction is defined as instruction where the instructor(s) and students are not required to meet at the same place or the same time.
- The use of FO instruction in mathematics is growing.
- Current research lacks clarity regarding the quality of learning outcomes and the influence of the disciplinary context.
3. The Nature of Mathematical Knowledge, Instruction and Assessment

Regarding teaching and learning, Freudenthal (1991) argued that mathematics is ‘learned differently and therefore should be taught differently’ (p.11). Regarding assessment practice, Webb (1993) further argued that the ‘nature of mathematics itself and pedagogical approaches for teaching mathematics warrant consideration of specific assessment techniques’ (p.662). In relation to these claims, this chapter reviews the literature on how mathematics is currently perceived, taught and assessed. First, further arguing for the need to attend to the disciplinary characteristics of mathematics, the chapter begins by presenting meta-analytic studies, comparing F2F to FO instruction, from the perspective of mathematics. Second, the nature of mathematical knowledge is discussed. Third, research on the current nature of mathematics teaching and assessment practice is reviewed. While some reflective literature is included, the emphasis will be on empirical findings, most of which investigates mathematics, or a related disciplinary grouping, in relation to other disciplines. In balance, the literature presents an overall picture of the hurdles and challenges facing current F2F instruction, which is contrasted in chapter four with what is known about current FO instruction.

3.1 Previous Meta-Analysis from a Mathematics Perspective

As discussed in the previous chapter, apart from some reasons given for scepticism, most of the research on FO instruction presents a fairly positive outlook. However, upon closer inspection, a different picture emerges from the disciplinary context of mathematics instruction. Three meta-analyses, including two previously discussed, are illustrative. First, in Bernard, Abrami, Lou, Borokhovski, Wade et al.’s (2004) study, while finding no significant difference overall, a closer investigation found that ‘math, science, and engineering appeared to be best suited to the classroom’ (p.400). Second, in Zhao et al.’s (2005) study, which also found no significant difference overall, only one study in mathematics met the inclusion criteria. This study examined achievement and post-secondary success of students in the same courses and found no difference in F2F and DE outcomes. However, these were students in a senior high school advanced mathematics course so one might expect students with above average ability in mathematics. Third, a recent meta-analysis of 126 qualifying distance learning studies (after strict inclusion criteria; examining 20 years of research studies and 1,850 comparative papers) includes only four studies in mathematics: two at the secondary level and two at the tertiary; with both tertiary studies presenting very mixed results (Shachar & Neumann, 2010).
summary, a pattern emerges that suggests greater uncertainty, compared to overall de-
disciplined study results, regarding the viability of distance teaching of mathematics.

In the context of e-learning, the lack of attention to the mathematics disciplinary context has
been an issue raised in the literature. Smith, Torres-Ayala and Heindel (2008), for example, in
their qualitative study comparing mathematics-related to non-mathematics related disciplines,
hypothesize that much of the current FO research is built on what they term ‘a tacit assumption
of...homogeneity’ and conclude that ‘the demands and solutions of e-learning’ are discipline-
specific and that disciplinary context is a ‘vital, yet largely overlooked, factor in research on e-
learning course design’ (p.82-83). And consistent with this hypothesis, Cretchley (2005) warns
against ‘interpreting and generalizing findings from other disciplines [into mathematics]’ (p.3).
As a first step to help address this lack of attention, this chapter now sets out to investigate how
mathematics currently differs from other disciplines.

3.2 Nature of Mathematical Knowledge

This section sets out to present a brief overview of the literature on the nature of mathematical
knowledge. This is accomplished by first considering some of the research that considers
mathematics in relation to other disciplines. This research is then shown to be consistent with
some of the reflective literature in mathematics education. Finally, a brief discussion on the
nature of mathematical communication further helps characterize mathematical knowledge. In
so doing, the presented characteristics are regarded as factors for consideration in pedagogical
development in general and innovation in particular (Lattuca & Stark, 1995).

At least three taxonomies have emerged, from both quantitative and more recently qualitative
research, which provide some insights into the ways mathematics is understood and experienced
in relation to other disciplines. These taxonomies classify academic disciplines into four
quadrants. For example, perhaps the most prominent work, Biglan (1973) classifies
mathematics as hard and pure, according to academics’ views. Hard (vs. soft) is meant to
convey that there is a high ‘degree of agreement about central theories, methods, techniques, and
research topics’ within mathematics (Smeby, 1996, p.70). Pure (vs. applied) is meant to convey
that mathematics is understood and experienced as being less concerned with practical
application. Consistent with Biglan’s findings, Kolb (1994) found students in hard pure
disciplines (e.g. mathematics) used an abstract (vs. ‘concrete’) and reflective (vs. ‘active’)
learning style. Finally, further consistent with previous findings Becher (2001) classifies
mathematics, as part of the hard pure family, as convergent and urban. The ‘convergent’ (vs. divergent) metaphor is meant to convey that academics in mathematics are generally in agreement and display a common or uniform understanding of the discipline. The ‘urban’ metaphor refers to a ‘close-knit’ epistemological structure and a competitive and densely populated research community. Despite some criticism (c.f. Kreber, 2009), these frameworks (Biglan’s and Becher’s in particular) have received fairly widespread acceptance and present a useful comparative means of contextualizing how mathematics is currently experienced.

These perspective characteristics are echoed and extended in the reflective literature. For example, characterizing mathematical knowledge as highly structured, Ernest (2010) emphasizes the interconnectedness, interdependence and ‘overall unity’ (p.47) of mathematical knowledge. And along with this sense of structure, Biberman-Shalev, Sabbagh, Resh, & Kramarski (2011) refer to a sense of immutability – mathematical knowledge as both ‘static and stable over time’ (p.8). Finally, in relation to other disciplines, and characteristics such as the latter, Azzouni (2007) argues that mathematics is ‘shockingly’ (p.204) distinct.

Finally, one more way mathematics may be differentiated from other disciplines is the nature of mathematical communication. First, mathematics uses a highly codified and well-defined symbol system (Smeby, 1996) which is considered to be both unique and complex (Quinnell & Carter, 2012). Second, as in the broader science and engineering fields, mathematics can be heavily dependent upon the use of diagrams and pictures to show relationships (Brown & Bakhtar, 1988). In short, where other disciplines communicate through the use of written text, mathematics relies heavily on notation, symbols and diagrams. And where text or writing-based soft disciplines rely on the Roman alphabet which is firmly embedded in the ubiquitous QWERTY, for mathematics there is no current equivalent ubiquitous input device. Instead, the ‘low-tech’ blackboard or pencil and paper, for example, still remains an effective means of communication for mathematics.

In summary, these characteristics of mathematical knowledge provide some idea of how mathematics is likely to be instructed and assessed. Furthermore, as will be discussed in chapter four, some of these characteristics present significant challenges to FO instruction.

### 3.3 Instruction and Assessment in Mathematics Courses

Instruction and assessment practice appear consistent with the nature of mathematical knowledge. That is, for example, the instruction and assessment of mathematics courses appears
to reflect an emphasis on transmitting information about its structure and connectedness, with assessment conducted by use of exams to see if this information has successfully been transferred.

### 3.3.1 Instruction in Mathematics Courses

The following section synthesizes findings related to instruction in undergraduate mathematics courses. Current findings indicate that, when compared to other disciplines, mathematics instructors and instruction are more ‘transmissive’ and least likely to emphasize ‘deep approaches to learning’.

Research suggests that teaching in the hard-pure disciplines is largely transmissive using mass lectures while soft disciplines tend to be discursive and involve more F2F contact (Neumann et al., 2002). First, Lueddeke (2003) surveyed 135 teaching staff from business, social science and technology using the Approaches to Teaching Inventory (ATI; Trigwell & Prosser, 2004). He found support for the hypothesis that the paradigmatic status of a discipline is an important factor influencing approaches to teaching; with paradigmatic disciplines (e.g. mathematics) found, for example, more likely to be transmissive, while pre-paradigmatic disciplines (e.g. English) were more likely to emphasize conceptual understanding. He further demonstrated that the discipline has one of the strongest influences on how teachers engage in ‘learning about’ and ‘demonstrat[ing]’ teaching knowledge. Second, more directly related to mathematics, findings from Lindblom-Ylänne, Trigwell, Nevgi and Ashwin (2006) confirm this contention. Their study of 340 Finnish and UK teachers used the ATI with Biglan’s framework. Among their findings, teachers from hard disciplines were ‘more likely to report a more teacher-focused approach to teaching, whereas those teaching ‘soft’ disciplines were more student-focused’ (p.294). Third, Cashin and Downey (1995) investigated faculty (and student) perceptions of objectives using a scale of 1-minor to 3-essential important. They surveyed 101,710 US college and university classes (including 5150 mathematics and statistics classes), across eight disciplinary groupings including one mathematics and statistics group. From a faculty perspective, the mathematics and statistics faculty were distinct from all other disciplinary groupings in viewing ‘learning to apply course material to improve rational thinking, problem-solving, and decision making’ as the most essential objective and ‘learning how professionals in this field go about the process of gaining new knowledge’ as the least essential (p.84). Additionally, providing more evidence of a predominantly transmissive mode of instruction, the objective of students gaining factual knowledge was weighted above average for mathematics
and statistics classes compared to the overall average of all disciplines. Fourth, Barnes, Bull, Campbell, and Perry (2001), surveying a national sample of 442 US faculty along 10 teaching goals, found that faculty in hard pure non-life disciplines (e.g. mathematics) were the least likely of all disciplines represented to prioritize development of creative thinking, management skills, leadership skills, commitment to one’s own values, respect for others and emotional well-being and only the most likely to prioritize learning of terms and facts - though analytic skills as a goal was just below faculty in soft pure life disciplines (e.g. psychology). Fifth, Nelson Laird et al.’s (2008) recent large scale study surveyed more than 80,000 students from 517 four-year colleges and universities and more than 10,000 faculty from 109 four-year colleges and universities in the US. They sought to examine ‘the effect of discipline on student use of and faculty members’ emphasis on deep approaches to learning as well as on the relationships between deep approaches to learning and selected educational outcomes’ (p.469). In their study ‘‘deep learning’ represents student engagement in approaches to learning that emphasize integration, synthesis, and reflection’ (p.469) and they use three sub-scale measures to quantify ‘deep approaches to learning’ across the disciplines – higher-order learning, integrative learning and reflective learning. Among their findings, of all disciplines, hard-pure-non-life faculty (e.g. mathematics) were the least likely to emphasize deep approaches to learning. In summary, when compared to other disciplines, current approaches to teaching mathematics tend to reflect an orientation to teaching that is teacher-focused and emphasizes transmitting facts and computational procedures.

### 3.3.2 Assessment Practices in Mathematics Courses

So far, this review has summarized areas of disciplinary difference with regards to the nature of mathematical knowledge, communication and instruction. While mathematics assessment practice will be discussed in a later chapter, this final section discusses assessment practice from the perspective of disciplinary differences.

What is known regarding disciplinary differences comes primarily from reflective papers. Perhaps the most commonly held notion is that hard disciplines employ problem-type examinations while soft disciplines employ essay-type examinations or term papers (Brown, 2004; Elton, 1998; Neumann, 2001). In particular, paradigmatic disciplines such as mathematics, with greater consensus on theories, are viewed as using more objective assessment approaches whereas pre-paradigmatic disciplines (e.g. soft), where there is less consensus on theories and methods, use more subjective approaches (Barnes et al., 2001). Similarly, hard
disciplines, such as mathematics, are viewed as emphasizing summative assessments with little (e.g. just a grade) or no feedback, quantitative questions and accuracy in grading whereas soft disciplines are seen as emphasizing a mix of ‘formative’ and ‘summative assessments’, qualitative questions and subjective judgments (Neumann, 2001; Neumann, Parry, & Becher, 2002). In summary, the literature identifies different assessment practices for different disciplines.

3.4 Summary

- Current meta-analytic findings across disciplines are much less clear when considered from the perspective of mathematics alone. Disciplinary differences are of interest in understanding the current thesis study context.
- The nature of mathematical knowledge is viewed as a distinct, unchanging and highly structured body of knowledge.
- Traditional undergraduate mathematics instruction is found to be largely ‘transmissive’ and the discipline least oriented to ‘deep approaches to learning’. Assessment practice in mathematics is viewed as using more objective and summative assessment instruments.
- Together, claims and findings provide some background that help frame challenges related to current FO mathematics course development.
4. Fully Online Mathematics

This chapter moves from the mathematics disciplinary context to discussing the characteristics of the FO instructional context. This is done so that current disciplinary characteristics of mathematics instruction can be compared with current characteristics of FO instruction. Out of this the idea emerges that mathematics instruction does not fit well with current dominant characteristics of FO instruction.

This chapter is structured in five sections. First, a distinction is made between instruction that is aided by and that which is immersed (i.e. FO) in computer-mediation. Second, the current state of research and development of FO mathematics is considered. Third, findings on rates of attrition are summarized. Fourth, the idea of ‘fit’ is introduced and related issues are considered. And fifth, the chapter concludes with a brief discussion on the quality of learning in FO mathematics courses. A major objective of this chapter is to provide a sense of the issues confronted as the cultures of tertiary mathematics and FO instruction meet. In so doing, this background context helps to consider the influences that affect the way assessment is practised in FO mathematics courses.

4.1 Web-Assisted vs. Web-Based Distinction

Mathematics has been recognized as a discipline with a greater affinity for computer-mediation, such as web-assisted instruction or assessment, than other disciplines (Fey, 1989; Laborde & Sträßer, 2010). However, it is unclear whether this affinity extends to more immersive forms of computer-mediation such as web-based or FO instruction. For example, it has been noted that computers are able to efficiently simulate, model, or help visualize mathematical concepts (e.g. computer algebra systems (CAS); Kemp & Jones, 2007; Smith, Torres-Ayala, & Heindel, 2008; White & Liccardi, 2007). Additionally, studies continue to associate the use of computer-assisted assessment (CAA) systems with mathematics assessment and most appear to do so in a positive light (e.g. Pitcher, Goldfinch, & Beevers, 2002). However, an important distinction is made in the literature between what has been termed web-assisted (structured around the lecture) vs. web-based (with no live lecture; e.g. ‘complete package’, Allen, 2003, p.270). For example, Laborde and Sträßer’s (2010) 25 year retrospective of ICMI activity, regarding technology use

---

3 CAA has been broadly defined as the use of ‘computers to deliver, mark or analyse assignments or exams’ (Sim, Holifield, & Brown, 2004, p.217). Given the initial computer-mediated state of FO instruction, this study extends the definition to include the use of computer automation in question generation and/or grading.
in mathematics, highlights a gulf between promise and actual practice. Related to the web-assisted vs. -based distinction, this gulf may be at its widest regarding the current viability of FO mathematics instruction. As this chapter will discuss, while web-assisted instruction shows practical potential, current web-based FO instruction presents a different set of issues that appear to call into question its viability.

4.2 Overview of Current FO Mathematics Research

Current research does not present a clear picture of the potential of FO mathematics instruction. Some see FO mathematics instruction as potentially superior to F2F instruction because of the benefits afforded by student anonymity. For example, it is claimed that anonymity enables students to feel less threatened by the prospect of confrontations that may commonly occur in live F2F settings (Mayes, 2004; Smith, Ferguson, & Caris, 2003) or able to adopt different online identities that enable ‘playful exploration’ (Rosa & Lerman, 2011). In short, some see the FO course context as enabling the learning of mathematics.

Other research supports a sceptical view of the potential of FO mathematics courses, that mathematics does not appear to fit well with FO provision (e.g. Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004). For example, since Kloeden and McDonald’s (1981) earlier paper outlining some persistent difficulties teaching mathematics at a distance, FO mathematics continues to be recognized as one of the most ‘difficult’ (Lokken, 2011) and ‘most challenging’ (Glass & Sue, 2008) disciplines to offer FO. While some research shows successful FO mathematics instruction is sensitive to such things as student characteristics (e.g. McIntosh & Morrison, 1974) and the nature of the course context (e.g. Zhao et al., 2005), the view that mathematics is challenging and difficult to teach in the FO context is echoed by a number of current reflective papers and empirical studies directed at FO mathematics instruction (Engelbrecht & Harding, 2005b; Glass & Sue, 2008; Philip, 2003; Smith, Ferguson, & Caris, 2003; Smith & Ferguson, 2005; Zinger, 2006).

As reflected in meta-analytic research, previously discussed, research comparing F2F and FO mathematics instruction is sparse and much of what does exist appears to suffer from similar methodological issues as the general FO research (Bernard, Abrami, Lou, & Borokhovski, 2004). Research on FO statistics instruction, for example, has shown mixed results – some consistent (e.g. Suanpang, Petocz, & Kalceff, 2004) and others inconsistent (e.g. Summers, Waigandt, & Whittaker, 2005) with equal or better results for the FO condition.
Results from large scale comparative studies also provide mixed results. For example, Mills and Raju (2011) conducted a 10 year review focused on statistics instruction analyzing 20 papers both reflective and empirical, many of which compared F2F to FO instruction. Despite acknowledging ‘a need for well-designed studies that control for confounding variables and other challenges related to empirical research’ (p.22) they appear to make little attempt to distinguish studies apart from their inclusion criteria (i.e. peer-reviewed studies). Despite also claiming FO statistics to ‘be a realistic option for years to come’ (p.21), one of their most stable findings across all studies reviewed is that students are dissatisfied with their FO statistics courses. This is consistent with other study findings concerning students’ experiences in FO mathematics courses (e.g. Summers et al., 2005). Second, Xu and Jaggars (2011) conducted a study focused on comparing F2F to FO provision in introductory college-level math and English courses by analyzing ‘administrative data’ (e.g. grades) from 24,000 students across 23 community colleges in one US state. Their results strongly suggest that for both disciplines FO instruction ‘at least as currently practiced, may not be as effective as F2F instruction at 2-year community colleges’ (p.374), with FO mathematics instruction singled out as faring more poorly than English. In balance, taking into account the methodological flaws in the first study, these findings are not consistent with equal or better instruction in FO (vs. F2F) courses. If anything, these studies support a sceptical view of the potential of FO mathematics instruction.

4.3 Attrition in FO Mathematics Courses

Current attrition statistics present perhaps the most convincing prima facie evidence in support of a sceptical view of the potential of FO mathematics. While it is commonly understood that rates of attrition in FO courses are generally higher than their F2F course counterparts (Lee & Choi, 2011; Merisotis & Phipps, 1999), at least three studies provide some perspective regarding attrition in FO mathematics courses. First, Mensch (2010), using data from over 14,000 student course enrolments at one US university, compared online mathematics and mathematics-related courses to other courses. He found significantly higher withdrawal rates for FO mathematics courses compared to other online courses and compared to F2F equivalent courses. Business statistics, for example, had a 32% FO versus 12% F2F withdrawal rate. This compared to an overall average withdrawal rate for all FO courses of about 10%. Second, Xu and Jaggars’ (2011) large scale study, previously discussed, found the attrition rate for introductory college math courses was 12% for F2F and more than double that (25%) for FO courses. Third, Smith and Ferguson (2005) conducted a two-part study surveying 138 State University of New York
Learning Network online instructors (including 32 from mathematics). They first compared FO mathematics to non-mathematics courses and second, using institutional data from their own institution, they compared F2F mathematics to non-mathematics courses. In the first study they found a significant difference in reported rates of attrition in FO mathematics courses (31%) vs. non-mathematics courses (18%) and, in the second study of just F2F courses, no significant difference (both at 5%). The clear suggestion is that FO mathematics students tend to be both quitting more than those students that are in FO non-mathematics as well as those that are in F2F mathematics courses.

4.4 Five Issues of ‘Fit’

There are at least a few possible reasons why the current FO course learning environment presents a challenging context for mathematics instruction. Many of these may be understood by contrasting the current disciplinary culture of mathematics, discussed in the last chapter, with elements of the current culture of FO instruction. In so doing the notion of ‘fit’ is presented with regards to issues of how these two cultures appear to clash. The following sections discuss five particular areas: four associated with characteristics of FO instruction and one associated with a characteristic of the FO medium.

Underlying many of these issues is the dominant pedagogical approach used in FO instruction which, as Anderson and Elloumi (2008) state, is ‘commonly based on constructivism’ (see also Keengwe & Kidd, 2010). This approach is consistent with the emphasis on discussion in FO courses which is perceived by many researchers to be one of the ‘most unique...sources of learning in online courses’ (Swan, 2003, p.25)

This thesis uses Richardson’s (2003) term, ‘constructivist pedagogy’, which she states constructivism, as a theory of learning, has ‘guided most of [its] development’ (p.1624). According to Richardson constructivist pedagogy is characterized by:

1. **Attention to the individual and respect for students’ background and developing understandings of and beliefs about elements of the domain (this could also be described as student-centered);**

2. **Facilitation of group dialogue that explores an element of the domain with the purpose of leading to the creation and shared understanding of a topic;**
3. Planned and often unplanned introduction of formal domain knowledge into the conversation through direct instruction, reference to text, exploration of a Web site, or some other means;

4. Provision of opportunities for students to determine, challenge, change or add to existing beliefs and understandings through engagement in tasks that are structured for this purpose; and

5. Development of students’ metawareness [sic] of their own understandings and learning processes. (p.1626)

4.4.1 Constructivist Pedagogy

The first of five issues of fit contrasts the dominance of constructivist pedagogy in FO instruction (Anderson & Elloumi, 2008) with its lack of emphasis in mathematics instruction. There are at least two reasons why constructivist pedagogy is not currently emphasized in mathematics instruction. First, it is well-documented that constructivist pedagogy is used infrequently in tertiary mathematics instruction (Neumann, 2001; Neumann, Parry, & Becher, 2002; Walczyk & Ramsey, 2003). Thus, for the many instructors that teach both F2F and FO, constructivist pedagogy may be an entirely unfamiliar or foreign experience. Second, some have noted that constructivism is incompatible with an objective view of reality (Bednar, Cunningham, Duffy, & Perry, 1995) or mathematical knowledge (Freudenthal, 1991). Therefore, assuming a dominant absolutist and objectivist view of mathematics, it would appear that constructivist pedagogy, at some level, would be in conflict or incompatible with the current disciplinary culture of mathematics. Though some contend that this incompatibility only applies to more extreme forms such as radical constructivism (Schoenfeld, 1992), what is happening in practice gives credence to the notion of a problem of fit. In short, while constructivist pedagogy appears to be emphasized in FO instruction (Anderson & Elloumi, 2008), its influence is seen to be ‘waning’ within mathematics (Confrey & Kazak, 2006, p.331).

4.4.2 Student-Led Learning

The second issue of fit is the expectation that students take a lead role in their learning (e.g. Anderson & Elloumi, 2008). This issue is as much a part of the structural nature of FO instruction (e.g. no regular class meetings) as an emphasis on constructivist pedagogy (Vrasidas, 2000).

In relation to this emphasis, much has been said about the nature of online instruction vis-a-vis the duelling metaphors: ‘guide on the side’ vs. ‘sage on the stage’ (e.g. Lim, 2004; Mazzolini &
Maddison, 2003). Similarly, the literature on FO mathematics instruction uses words describing the instructor’s role as ‘mediator’ (Maltempi & Malheiros, 2010), ‘coach’ (Evia, 2006) or ‘facilitator’ (Smith, Ferguson, & Caris, 2003).

However, while there are those that advocate and see the potential for mathematics students to take more responsibility for their own learning (e.g. Houston, 2002), it is clear that current F2F practice, by and large, does not reflect such a reality. As presented in the previous chapter, current mathematics instruction tends to be teacher not student-led (Lindblom-Ylänne, Trigwell, Nevgi, & Ashwin, 2006; Pampaka et al., 2011), with a disciplinary culture that tends to engender dependence and not the independence requisite for success in FO instruction (Kerr, Rynearson, & Kerr, 2006). Consistent with such a view, White and Liccardi (2007) found, when comparing soft vs. hard disciplines in e-learning, that students in hard disciplines valued passive and teacher-led activities, while students in the soft areas valued active and student-led activities – concluding that e-learning is more suited for some disciplines than others. In this respect, it seems understandable why many find teaching mathematics FO to be ‘exceedingly awkward’ (Smith, Torres-Ayala et al., 2008, p.74).

4.4.3 Use of Discussion

A third issue of fit is the emphasis on online discussion (e.g. Swan, 2003), where its use is typically realized as a weighted part of an FO courses overall assessment scheme (Everson & Garfield, 2008; Smith, Ferguson, & Caris, 2003). There are at least two reasons why discussion does not fit. First, there are problems with communicating mathematics in online discussions, which will be discussed in a later section. Second, compared to other disciplines, researchers have claimed (e.g. Entwistle, 2009) that some disciplines, such as mathematics, are less likely to use discussion than others.

However, many claim online discussion in FO mathematics courses has potential. For example, the use of discussion in FO mathematics has been linked to the development of a sense of community (Everson & Garfield, 2008; Gadanidis, Graham, McDougall, & Roulet, 2002) and to a better quality of learning (Engelbrecht & Harding, 2005b; Mallet, 2008; Offenholley, 2006). Though, despite such potential, the current research on online discussion in mathematics presents a mix of opposing claims and empirical findings which appear limited by a failure to define and distinguish different forms of discussion. Pirie and Schwarzenberger (1988), for example, define discussion as ‘purposeful talk on a mathematical subject in which there are
genuine [student] contributions and interaction’ (p.461). However, much of the research fails to
distinguish this kind of discussion from, for example, a discussion reply that is simply a
statement of agreement or discussion concerning assessment due dates. With this limitation in
mind, some of the current findings are summarized, first those against and then in favour of the
potential of FO discussion in mathematics.

First, the strongest findings suggest, at the very least, that the use of discussion is not perceived
to be beneficial. For example, Smith, Heindel and Torres-Ayala (2008), using log files from
over 500 students in two semesters separated by five years, studied transactional distance
(Moore, 2007) across Biglan’s four disciplinary groupings. In brief, Moore’s (2007) theory of
transactional distance viewed student dialogue, course structure and student autonomy as three
pedagogical variables that mediate for what has been described as the perceived sense of
isolation a student feels. Smith et al. found that for students in pure online courses, the
transactional distance (i.e. sense of isolation) was widening relative to students in applied
courses, where this distance was shortening. In their conclusion, they refer to the
‘commoditization’ of pure online courses (versus the community orientation of applied). In
another study, Finnegan, Morris and Lee (2009) analyzed log files and academic records of 118
sections of 22 FO courses taught at six University System of Georgia institutions comprising
grades from approximately 2500 students. Of the three disciplinary groups analyzed, they found
students in ‘science, technology and math’ had the highest mean viewing time of ‘content pages’
(5.87 hrs compared to 3.19 for English and Communication and 5.32 for Social Sciences) and
less than half the mean viewing time reading ‘discussions’ of the other two groupings (2.21 hrs
compared to 5.43 and 6.01). They conclude, from a multiple regression analysis of discussion
variables across achievement, that active STEM student participation in discussion was one of
the least worthwhile investments in yielding better grades. In another, Illowsky (2007)
conducted a study at two US community colleges where students who self-enrolled in an FO
elementary statistics course were randomly placed into a control (n=60 total) and an
experimental group (n=46 total) where the ‘intervention in the experimental groups was required
discussion postings involving higher order critical thinking skill’ (p.61). She found that the ‘use
of structured higher order thinking online discussion postings in online elementary statistics
classes’ did not contribute to ‘higher academic success rates’ (p. i). Finally, some recent
qualitative research, at the school-level in Australia, found teaching of mathematics at a distance
is a primarily ‘one-on-one’ experience with little use of collaboration (Lowrie & Jorgensen,
2012). In summary, at least among students, as compared to other disciplines, discussion does
not appear to be emphasized, with some evidence that it is not helping improve student performance.

Second, mostly from a weaker basis, only a few studies find in favour of the use of discussion in FO mathematics courses. In one study, Gorsky, Caspi, Antonovsky, Blau, and Mansur (2010) conduct a mixed method log file and content analysis of 50 online forums comprised of 4903 students in ‘exact and natural sciences’ and humanities (i.e. hard vs. soft). They found that active participation was much higher in the sciences than the humanities, contending that this was because science courses are generally more challenging than humanities courses and thus require more interpersonal dialogue. However, apart from the broader disciplinary focus, given the Open University context, there is some question regarding, for example, possible effects of student demographic characteristics (e.g. mature students).

In summary, the balance of research appears to support a problem of fit with the use of discussion in FO mathematics courses. How much this is due to the dominant transmissive approach used in F2F mathematics instruction (e.g. Barnes et al., 2001) is unclear. However, given discussion is linked to gaining understanding in mathematics (e.g. Skemp, 1979), its use will be a significant focus of the thesis study.

4.4.4 Use of E-Lectures

As a fourth issue, the use of e-lectures may be prevalent (e.g. PowerPoint slides, recorded ‘live’ lectures) given there are no live classes or lectures and perhaps particularly so given that the use of discussion does not appear to work well as an instructional tool. However, early research in this area presents some questions about the value of e-lectures.

Trenholm, Alcock and Robinson (2012) recently reviewed the mathematics e-lecture research and found, while e-lectures were positively received by instructors and students, ‘findings appear to indicate cause for concern regarding the implementation and use of e-lectures in mathematics’ (p.6). In particular, the research that was reviewed indicated that the use of e-lectures was correlated with lower grades and, as suggested by one study, to be ‘enabling’ surface learning (Le, Joordens, Chrysostomou and Grinnell, 2010, p.5).

On balance, the instructional value of mathematics e-lectures remains unclear. And though positively received by students this may refer to their use as a supplement rather than a replacement for F2F lectures. As White & Liccardi (2007) found, compared to students in the
soft disciplines, those in the hard disciplines (e.g. mathematics) expressed a desire not to abandon the traditional F2F lecture.

### 4.4.5 Communicating Mathematics

A final issue of fit relates to current problems with communicating mathematics in the FO course context. There is clear recognition that problems exist and persist with communicating mathematics in the FO context (Gadanidis et al., 2002; Lingefjärd & Holmquist, 2002; Philip, 2003; Smith, Torres-Ayala et al., 2008). Of all the issues of fit discussed, this is perhaps the most serious given its potential impact on hindering the full expression or reality of the mathematics disciplinary culture (Kramsch, 1998, Becher & Trowler, 2001).

Evidence of these problems is apparent when discussing the two composite elements needed for online communication: computer hardware and software. First, while the qwerty-keyboard and mouse present a relatively natural fit with, for example, text-based soft disciplines, the same cannot be said for the heavily symbolic and diagrammatic nature of mathematics. Instructors may, for example, create their own notational systems (e.g. using $\text{sqr}(x^2-y)$ for $\sqrt{x^2 - y}$), scan and email, draw with a mouse or use tool palettes. While the latter three are obvious concessions to efficient and natural communication, the former is seen as enforcing ‘rigid syntax constraints... [which are]...difficult for students to learn’ (Smith, Torres-Ayala et al., 2008, p.72). Second, the potential of software such as what would be used in web-assisted instruction (e.g. web applets), appears limited to specific contexts or concepts. Outside of specific software the ability to freely and dynamically present mathematics (e.g. communicating diagrams, graphical representations or worked examples) may be seriously impeded or even non-existent (Smith, Torres-Ayala et al., 2008). The situation is such that a FO mathematics course may require, for example, the use of three websites – for example, use of interactive software (e.g. GeoGebra), a textbook website with CAA and a separate virtual learning environment\(^4\) (VLE) each with their own toolbar/palettes and/or key commands (Engelbrecht & Harding, 2005b). In short, writing and communicating mathematics in this context may be challenging and, compared to other disciplines, FO mathematics students may be more disadvantaged by the additional learning curve and the increase in ‘extraneous cognitive load’ involved in learning to effectively use such communication – further taxing what may already be a challenging disciplinary area for students

---

\(^4\) One could particularly single out VLE’s. While attempting to present a common framework for students to receive multi-disciplinary online instruction, they may fail to address the distinctive requirements of disciplines such as mathematics.
(Smith, Torres-Ayala et al., 2008). One consequence of which, discussed earlier, may be high dropout rates in FO courses which have been attributed to cognitive overload in the early stages of a course (Tyler-Smith, 2006).

In summary, there is some irony here. While mathematics has been considered the ‘language of computers’ (Willinger & Paxson, 1998, p.961), computers in the FO context currently appear to be impairing the ability of communicating the language of mathematics. Moreover, given many in FO instructional development view the establishment of community as ‘critical to the success of online courses’ (Swan, 2001, p.327), it is perhaps this issue coupled with the use of discussion that may challenge current FO mathematics instruction the most.

Together these five issues of fit provide some background regarding the nature of the current FO course context, particularly in relation to the disciplinary characteristics of current mathematics instruction. They provide some reasons why, overall, current FO mathematics courses do not appear to be working well. In this thesis, R1 of the first study will investigate how discussion, as a weighted assessment instrument, is being used in FO mathematics courses. R1 of the second study then seeks to explain how and why discussion and interactions, in general, are being used in these courses.

4.5 The Quality of Learning in the FO Instructional Context

Apart from and related to these problems of fit, there are two competing claims regarding the quality of learning in general FO instruction. First, that the internet medium, where FO courses reside, discourages higher quality learning. Second, that FO instruction, occurring within that medium, encourages higher quality learning. This section outlines both of these arguments.

E-learning in general (i.e. ‘high-level cognitive skills’, Laurillard, 2002) and FO instruction in particular, have been linked to higher quality learning (i.e. ‘conceptual learning’, Parker & Gemino, 2001). Regarding FO instruction, the argument set forth is at least three-fold. First, the typically asynchronous nature of FO instruction provides time for students to reflect and formulate their thoughts (Havard, Du, & Olinzock, 2005; Swan, 2001). As Swan (2001) states, FO instruction engenders a ‘culture of reflection’ among students (p.310). With regards to mathematics, this would seem particularly advantageous given reflectivity has been linked to the development of conceptual understanding (Engelbrecht, Harding, & Potgieter, 2005) and advanced mathematical thinking (Tall & Dubinsky, 1996). Second, FO instruction requires students to write out their thoughts and, as Caris, Ferguson, & Gordon (2002) found, instructors
view ‘the realization that those [i.e. students’] thoughts will be exposed semi-permanently to others in the class seem to result in a deeper level of discourse’ (p.65). This concurs with mathematics-specific findings where writing has been found to benefit ‘students’ construction of mathematical knowledge and assist in building higher levels of understanding’ (Miller, 2007, p.441). Third, the emphasis on discussions and interactions (In general F2F Kember & Gow, 1994; In FO mathematics Overbaugh & Nickel, 2011) and constructivist pedagogy (Ally, 2004), as previously mentioned, have been linked to higher quality learning. For example, with regards to mathematics, the potential of peer instruction in mathematics has been linked to improvements in the quality of student learning (Crouch & Mazur, 2001; Dochy, Segers, & Sluijsmans, 1999; Springer, Stanne, & Donovan, 1999).

However, many concerns have been raised in the literature with regards to the wider internet medium. Many of these relate to the central thesis of Carr’s (2010) recent book, *The Shallows: How the Internet is Changing the Way We Think, Read and Remember*. In his book, Carr argues that instruction mediated by the internet is instruction taking place in what writer Cory Doctorow terms an ‘ecosystem of interruption technologies’ (p.91) – which is essentially antithetical to higher quality learning. In mathematics education, a few examples appear illustrative. First, Gadanidis, Graham, McDougall and Roulet (2002) argue that students need to be ‘re-trained’ (p.12) in order to make the transition from seeing computers as tools for immediate gratification to tools for thinking mathematically. Second, Engelbrecht and Harding (2005b), referring to Stiles (2000), contrast multi-media as a tool to ‘excite with its ‘richness’ (p.3) versus for its value for learning. Third, as mentioned earlier, FO mathematics instruction may require task switching involved with using multiple websites and user interfaces (Engelbrecht & Harding, 2005b) where such multi-tasking may impede on learning (Smith, Torres-Ayala et al., 2008). As the learning environment is seen to influence ‘the study approach which students adopt’, one may expect some effect (Kember, 1996, p.352).

Whether this interaction, between the wider internet medium and the FO instructional context, supports higher quality learning is still unclear. For example, outside of mathematics, some have found FO students did better than F2F students on conceptual questions but not on questions about ‘technique’ (Parker & Gemino, 2001). In contrast, others found outcomes were ‘comparable’ at ‘lower levels of abstraction (recollection) but as work progresses to higher orders of abstraction (application) significant performance differences arise that place on-line learners at a disadvantage’ (Ross & Bell, 2007, p.3). Within mathematics the central interest of the present research relates to the investigation of the current nature of FO instruction and
assessment practice within this medium. In relation to the issue fit, R4 of the first study will investigate current professional development opportunities available and being used by FO mathematics instructors, particularly in relation to any focus on the disciplinary characteristics FO mathematics.

4.6 Summary

- FO mathematics instruction presents an interesting conundrum: The literature suggests it is well-suited for some level of computer-mediated instruction but not fully mediated instruction.
- Findings on rates of attrition suggest a problem of fit.
- Problems of ‘fit’ are manifested where current cultural norms in mathematics instruction (e.g. most instruction tends to be teacher-led and constructivist pedagogy is used infrequently) appear to conflict with norms in FO instruction (e.g. most instruction tends to be student-led and constructivist pedagogy is emphasized) and the affordances of communicating in the FO course context present significant challenges for mathematics instruction.
- In light of these issues, the chapter argues that the FO context presents, at best, a challenging environment for delivering quality FO mathematics instruction.
5. Research on Approaches to Teaching and Assessment

Having summarized some of background literature related to FO mathematics instruction, the remaining two chapters summarize the literature related to the focus of the thesis research: approaches to teaching and assessment followed by tertiary mathematics assessment practice. To investigate the nature of current FO mathematics assessment practice, this thesis study focuses upon instructors’ approaches, otherwise known as orientations (Hativa & Birenbaum, 2000) to assessment practice. In higher education, there has been an expanding body of research devoted to the study of approaches from a student’s as well as an instructor’s viewpoint (Richardson, 2005). The chapter begins by providing a brief overview, including some criticism, of the early theoretical work underlying current research on approaches. This work is then related to the literature on the quality of learning in mathematics. Next, framing the remainder of the chapter, research on approaches to teaching and assessment is related to the quality of student learning. This is followed by a summary of the literature on instructors’ approaches to teaching, in general, and in mathematics, in particular. Following this, the literature is summarized on instructors’ approaches to assessment in general and then to setting mathematics questions. Finally, two study frameworks selected for use in the thesis research are introduced and summarized.

5.1 Broader Theoretical Background: Deep Learning

In this section deep learning is introduced as it relates to the theoretical framework originating with Marton and Säljö’s (1976a, 1976b) early work. Following this, an attempt is made to map deep learning onto current research on the nature of understanding and learning in mathematics.

5.1.1 History

In the mid-1970’s, a new method was sought to collect and analyze data that would describe qualitative differences in learning outcomes – this gave rise to the study of ‘approaches to learning’ (Marton and Säljö, 1976a). As such, phenomenography was developed as a research methodology and findings indicated that students typically approach studying using either a deep or surface learning strategy. Marton and Säljö (1976a) explicate the deep/surface approach.

A third orientation, ‘strategic’ approaches to learning, was later added (Entwistle & Tait, 1990). This orientation is not discussed in this chapter as it relates to student and not instructor approaches.
metaphor: deep learning is learning ‘directed towards comprehending’ and referred to as ‘the signified’ and surface learning is a ‘reproductive’ conception of learning’ whose focus is on the symbols or text and referred to as the sign’ (p.7-8). Entwistle (2000) summarizes:

Interviews with adults who had different educational backgrounds enabled Säljö (1979) to identify a hierarchy of distinct conceptions of learning. The simplest conception saw learning as the accretion of discrete pieces of information into knowledge. In contrast, the most complete conception focused on learning as the development of personal understanding... (p.5)

Though criticized as such (Webb, 1997), it is important to note that these findings do not represent a simple dichotomy but, as Entwistle (2000) notes, a hierarchy of conceptions.

Recent work has moved beyond approaches to learning to approaches to teaching (e.g. Even & Kvatinsky, 2009) and more recently approaches to assessment (e.g. Samuelowicz & Bain, 2002). While these studies may not make specific reference to deep learning, they are conceptually similar in that they seek qualitative measures of how, for example, instructors view and for what purpose they use assessment instruments in their course.

Marton and Säljö’s framework has had an intuitive appeal. In the case of students, it considers how they approach their studying or learning. In the case of teachers, it considers how they approach their teaching or assessment practice. Such an appeal appears to have helped this framework gradually take root, particularly in the European educational research field (Richardson, 1999; Webb, 1997).

5.1.2 The Quality of Learning in the Mathematics Instructional Context

This section considers the quality of learning as it specifically pertains to mathematics. As this research remains focused on the instructor’s role in the student learning dynamic, this section provides some idea of what instructors view as the target of their instruction.

Within the mathematics education literature several have attempted to define the quality of learning, with some using the terms deep or surface learning. First, Skemp (1979) refers to three different qualities of understanding in mathematics: instrumental, relational and logical. Second, Gray & Tall (1994) identify ‘the successful mathematical thinker’ as someone that uses a ‘mental structure which is an amalgam of process and concept’ – termed a ‘procept’– with those less successful limited to procedural knowledge (p.115). Third, Smith and Wood (2000) define a surface learning approach as one that ‘concentrates on the learning of unrelated facts
and patterns, while a deep approach focuses on the relationship of the facts to the concepts’ (p.125). Fourth, Houston (2002) characterizes deep learning in mathematics as: ‘learning which is consolidated, learning which will be retained because it connects with previous learning, learning which develops curiosity and a thirst for more, learning which is demonstrably useful in working life’ (p.408). Fifth, Engelbrecht, Harding and Potgieter (2005) refer to perhaps the best known (Star, 2005) contrasting description of quality learning: ‘procedural’ vs. ‘conceptual knowledge’. They define procedural knowledge as ‘the ability to physically solve a problem through the manipulation of mathematical skills, such as procedures, rules, formulae, algorithms and symbols used in mathematics’ and conceptual knowledge as ‘the ability to show understanding of mathematical concepts by being able to interpret and apply them correctly to a variety of situations as well as the ability to translate these concepts between verbal statements and their equivalent mathematical expressions... a connected network in which linking relationships is as prominent as the separate bits of information’ (p.704,5). Sixth, Star (2005) further extends the previous terms to consider procedural and conceptual knowledge that is either ‘superficial’ or ‘deep’ (e.g. deep procedural knowledge). Star also discusses the idea of ‘flexibility’ as ‘an indicator of deep procedural knowledge’ (p.409). This is an idea Baroody, Feil and Johnson (2007) extend and refer to as entailing ‘flexible (as well as efficient and appropriate) application of procedures...that...both benefits from and benefits conceptual understanding’ (p.120). In summary, for the most part, these descriptions appear conceptually aligned with Marton and Säljö’s (1976a) framework. In particular, related to mathematics instruction and assessment practices, they suggest the different kinds of understanding possible as well as the kind of understanding instructors are likely to be targeting for development.

5.2 Instructors’ Approaches

5.2.1 Importance of Investigating Instructors’ Approaches

The importance of this investigation relates to studies that have generally shown that teachers’ approaches to teaching are related to students’ approaches to learning (Kember & Gow, 1994; Trigwell, Prosser, & Waterhouse, 1999). For example, Kember and Gow (1994) conclude that ‘where the knowledge transmission orientation predominates, the curriculum design and teaching methods are more likely to have undesirable influences on the learning approaches of students’ (and vice-versa for what they term a ‘learning facilitation’ orientation). Furthermore, it has been shown that ‘students who perceive the nature of the assessment as encouraging
memorisation and recall, and who perceive the workload demands of a subject as high, are more likely to adopt a surface approach’ (Ramsden, 2003 as cited in Trigwell et al., 1999, p.58).

Moreover, students’ approaches to learning have been shown to be related to higher quality learning outcomes (Trigwell, Prosser, Ramsden, & Martin, 1998). Considering this chain of implications, this study seeks to investigate instructors’ approaches to teaching and assessment with a view to gaining some understanding regarding the likely quality of related student learning outcomes.

5.2.2 Approaches to Teaching in General

Early research into instructors’ approaches to teaching has led to the development of a psychometric instrument. Trigwell, Prosser and Taylor’s (1994) study is often cited as one of the earliest studies which eventually led to the development of the Approaches to Teaching Inventory (ATI; Trigwell & Prosser, 2004). Since the ATI’s development it has received fairly widespread acceptance. The following section provides some background on the ATI.

Trigwell et al.’s (1994) early study sought to investigate university science teachers’ approaches to teaching. In this study they interviewed twenty-four university lecturers. Using a phenomenographic methodology they found five qualitatively different approaches to teaching, of which the two extreme categories (Trigwell & Prosser, 2004) became the focus for the development of the ATI. These categories were ‘teacher-focused strategy with the intention of transmitting information to students’ and ‘student-focused strategy aimed at students changing their conceptions’. From this approaches were ‘constituted’ in terms of strategies that teachers adopt and ‘the intentions underlying the strategies’ (p.413) – where intentions ranged from transmitting information to seeking conceptual change and strategies ranged from teacher- to student-focused. As such this was seen to present a means of describing teachers’ approaches to teaching.

From this basis, the ATI was developed as a psychometric instrument used to measure the ‘variation between an information transmission/teacher-focused view of teaching and a conceptual change/student-focused view of teaching’ (abbreviated as ITTF and CCSF, p.415). This was accomplished through an iterative process of selection and reduction where a 39-item version inventory was developed and trial tested on a sample of 58 university physics and chemistry teachers. Second, a principal components factor analysis and Cronbach’s test of reliability was run through which the inventory was reduced to 16 items with four subscales (see 53
Appendix A). In its final form, each of the ATI questions provided responses on a 5-point scale from only rarely true to almost always true.

The ATI was later revised (Trigwell, Prosser, & Ginns, 2005) to include 22 questions and an extensive confirmatory factor analysis was successfully conducted (Prosser & Trigwell, 2006). Despite some fierce criticism (Meyer & Eley, 2006), the ATI has generally been found to be a reliable and valid psychometric instrument (e.g. Stes, De Maeyer, & Van Petegem, 2010). R6 of the first study will use the ATI to investigate FO mathematics instructors’ approaches to teaching their FO courses.

5.2.3 Approaches to Teaching Mathematics

While there have been studies at the K-12 school level (e.g. Even & Kvatinsky, 2009; Kuhs & Ball, 1986) little is currently known about mathematics instructors approaches to teaching at the tertiary level. As Speer, Smith III and Horvath (2010) confirm: ‘at the collegiate level…very little empirical research has yet described and analyzed the practices of teachers of mathematics’ (p.99). In one of the few studies, at the further education level6, Pampaka, Williams, Hutcheson, Wake, Black, Davis and Hernandez-Martinez (2011), found that, consistent with tertiary findings discussed in the chapter on disciplinary differences and distinctiveness, most teaching is transmissionist and teacher-centred. Contrary to expectations, they found no relationship between pedagogic practice and student grades. They speculated this may be the result of student predispositions to ‘test-centeredness’ or surface learning where teachers’ approaches are ‘only one part’ of changing the overall student learning experience. Such findings appear to highlight the complex dynamics in mathematics instruction and assessment practice.

5.2.4 Approaches to Assessment in General

As with mathematics instructors’ approaches to teaching, little is known about their approaches to assessment (Speer et al., 2010). However, two studies were found with a focus on general F2F tertiary level assessment practice. First, Guthrie (1992) interviewed 239 above average faculty members from a wide array of disciplines in three US higher education institutions. These were faculty who, when compared to other faculty and courses, ‘produced students who demonstrated the highest gains in cognitive abilities, particularly analytical reasoning’ (p.69).

---

6 Further education, in the UK context, may be considered as roughly equivalent to ‘community colleges’ in the US context.
Guthrie was interested in understanding what these professors did that made them better than others at teaching – specifically, what were their goals, methods of instruction and approaches to final evaluation. Regarding approaches to final evaluation, interview questions were developed to address three domains: cognitive, affective and psychomotor with the former, cognitive domain, developed based on Bloom’s (1956) taxonomy. Among the main findings, faculty were found to primarily emphasize cognitive development. And despite class interaction being valued, little assessment weighting was assigned to it. Finally, despite producing students with superior cognitive gains, it was found that final evaluations (e.g. final exam) generally emphasized ‘lower-order cognitive abilities’ (p.76). Studies like this further highlight some of the complexity involved in investigating FO assessment practice. In particular, as F2F instructors who meet at a regular time and place with their students, it is perhaps understandable that class interaction may be valued and used while receiving little assessment weighting. However, for FO instructors without a regular time and place to meet with their students, it is perhaps understandable that these interactions are more likely to be incentivized through assigned assessment weighting.

Second, Samuelowicz and Bain (2002), seeking to identify academics’ orientations to assessment practice, interviewed 20 academics from a mix of disciplines at three different Australian universities. Using a grounded theory approach they identified six global orientations to assessment (that could be ordered along a continuum from knowledge reproduction to knowledge construction and/or transformation), as well as six qualitative belief dimensions. Table 1 presents an overview of the S&B findings with orientations as rows and beliefs as columns. This framework identifies the nature of assessment practice and bears ‘broad similarities’ (p.194) to previous research related to deep learning. There also appear to be some obvious similarities with the ATI. For these reasons, this framework was selected for investigating the approaches to assessment of FO mathematics instructors.
Table 1: Framework presenting S&B study findings

<table>
<thead>
<tr>
<th>Orientations to assessment practice (with associated ranks)</th>
<th>Belief Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nature &amp; structure of knowledge - Q10</td>
<td>Academic believes knowledge to be assessed as...</td>
</tr>
<tr>
<td>Academic views knowledge to be assessed as...</td>
<td>2. Degree of integration of knowledge - Q11</td>
</tr>
<tr>
<td>Academic believes that assessments should...</td>
<td>3. Degree of transformation of knowledge - Q12</td>
</tr>
<tr>
<td>Academic believes that assessments should...</td>
<td>4. Differences between good &amp; poor answers - Q13</td>
</tr>
<tr>
<td>5. Role of assessment in teaching &amp; learning – Q14</td>
<td>Academic believes that feedback from student performance should be used to...</td>
</tr>
<tr>
<td>6. Use of feedback gained from assessment – Q15</td>
<td>...alter his/her teaching.</td>
</tr>
<tr>
<td>1: Reproducing Bits of knowledge</td>
<td>...external to students and as a collection of important bits (definitions, concepts, techniques, methods, theories).</td>
</tr>
<tr>
<td>...draw on information presented in a single lecture, tutorial, practical session or chapter.</td>
<td>...determine whether students can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes.</td>
</tr>
<tr>
<td>2: Reproducing structured knowledge</td>
<td>...require students to draw on information presented in many sources, but within their subject.</td>
</tr>
<tr>
<td>3: Applying structured Knowledge</td>
<td>...require the application of well known techniques, methods, laws, principles, or explanations to unseen standard problems.</td>
</tr>
</tbody>
</table>

| 4: Organising subject knowledge                             | ...what has been internalised, reorganised, and reconstituted in the process of learning. |
| ...require students to apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems. | ...assessment to be an integral part of teaching and learning, a means of helping students learn. | ...monitor students’ learning and to help them improve. | ...challenge students’ existing ideas and understandings. |
| 5: Transforming discipline knowledge                        | ...require students to integrate information from many sources, and/or from more than one subject, and their own experience. |
| ...assessment to be an integral part of teaching and learning, a means of helping students learn. | ...monitor students’ learning and to help them improve. | ...challenge students’ existing ideas and understandings. |
| 6: Transforming conceptions of the discipline/world         | ...require students to apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems. | ...assessment to be an integral part of teaching and learning, a means of helping students learn. | ...monitor students’ learning and to help them improve. | ...challenge students’ existing ideas and understandings. |
Samuelowicz and Bain (hereafter ‘S&B’, 2002) claim their study to be ‘the first attempt to provide a systematic description of the differences between academics’ orientations to assessment practice’ (p.194). Findings from their earlier work on orientations to teaching and learning (Samuelowicz & Bain, 2001) were found to be strongly related to orientations to assessment. Specifically, they found that teachers who emphasized ‘teaching as exposition and learning as reproduction tend to believe that assessments should determine how well students can reproduce the knowledge they have been given and how well they can use that knowledge in much-practised tasks… (whereas) those who view teaching as facilitating learning, and learning as constructing a personal understanding based upon established knowledge and procedure, tend to believe that assessments should require purposeful transformation of knowledge to address open-ended issues or problems not previously encountered’ (p.196). Their main findings were seen to confirm that ‘orientations to assessment practice would range from those favouring the reproduction of knowledge and procedure to others favouring the construction and/or transformation of knowledge’ (p.197). Since its publication, the study has been cited numerous times in subsequent research. However, this appears to be the first time it will be used as a framework for analyzing specific instructor assessment practices. R5 of the first study will use the S&B study as a framework for investigating FO mathematics instructors’ approaches to assessment in their FO courses. R5 of the second study then seeks to gain a more elaborate understanding of these findings using interviews.

5.2.5 The ATI compared to the Samuelowicz and Bain 2002 Study

The ATI presents an instrument, accepted within the education research community, to help investigate the nature of FO mathematics instructors’ teaching practices. The S&B study presents a viable framework to help investigate the nature of FO mathematics instructors’ assessment practices. In so far as they will be used concurrently in the thesis study, similarities and differences between them should be noted. Framing any discussion are problems concerning the use of terminology in the two studies, problems that have been noted in the wider literature on teaching ‘approaches’, ‘conceptions’ and ‘beliefs’ (Kember, 1997). For example, Trigwell and Prosser (1996) define an approach as comprising of an intention or motive and a strategy. S&B, on the other hand, use the term orientation, which others have viewed as

---

7 “By ‘orientation’ the authors mean “a coherent pattern of beliefs inferred from, and grounded in, academics’ assessment practices and their explanations of those practices” (p.176). This notion appears to be very similar to Trigwell and Prosser’s (1996) division of approaches into two components: intention and strategy.
functionally equivalent in meaning to approach (Hativa & Birenbaum, 2000). S&B defines orientation as ‘a coherent pattern of beliefs inferred from, and grounded in, academics’ assessment practices and their explanations of those practices’ and state that their study seeks to ‘probe for characteristic perspectives and practices’ (Samuelowicz & Bain, 2002, p.176).

Furthermore, the terminology and questions used in both studies provides further means of comparison. For example, in the ATI questions about ‘intentions’ are premised with ‘I feel’, ‘I encourage’ or ‘I think’ whereas statements of ‘strategies’ use statements that reflect actual practice such as ‘I use’ or ‘I teach’. S&B framework findings are presented as statements of ‘belief’ regarding ‘the nature and function of their assessments’ (p. 173) and premised with: ‘The Academic believes that...’ In balance, though problems with terminology persist, it appears reasonable to argue that both studies concern themselves with a similar underlying construct that combines intended and actual practice.

Finally, in terms of differences, the S&B study seeks, by using a grounded theory approach, instructors’ ‘characteristic perspectives and practices’ (italics mine; p.176). In contrast, in the ATI study, a phenomenographic research methodology is used that seeks variation in instructors’ approaches. However, in the end, the S&B orientations appear conceptually aligned with the ATI approaches – for example, the ATI CCSF approach with the S&B ‘knowledge construction’ orientation. R7 of the first study will use the ATI, as an established psychometric instrument, to investigate the validity of the S&B framework.

5.2.6 Approaches to Assessment in Relation to Mathematics Questions

Finally, while little is known about mathematics instructors’ approaches to assessment, some related work, largely based on Bloom’s (1956) taxonomy, has been done on classifying mathematics questions. First, Smith, Wood, Coupland, Stephenson, Crawford and Ball (1996), for a general instructional context, proposed a taxonomy to classify mathematics questions. Using Ramsden’s (1992) framework and Bloom’s taxonomy they constructed seven levels of questions from those requiring a surface to deep approach to learning. Second, Vidakovic, Bevis, and Alexander (2003), for the online CAA context, also used Bloom’s taxonomy to classify questions. Third, Pointon and Sangwin (2003), for the portable CAS context, again also used Bloom’s taxonomy to develop a ‘question classification scheme’ (p.675) with eight levels. These different classifications (see Table 2) appear aligned with the S&B orientations to
assessment, suggesting the use of the S&B framework has some relevance for studying mathematics assessment practice.

Other work analyzing mathematics textbook exercises and exam questions reflects an alignment with the extremes of the S&B framework. Bergqvist (2007), for example, investigated 16 Calculus exams from four different Swedish universities using Lithner’s (2000) framework that broadly considers two types of reasoning – creative mathematically founded reasoning and imitative reasoning founded on ‘recalling answers or remembering algorithms’. These bear strong resemblance to the S&B extremes of knowledge construction and knowledge reproduction. Furthermore, consistent with an orientation to teaching that emphasizes transmitting facts and computational procedures (see 3.3.1), she found that ‘about 70% of the [exam] tasks were solvable by imitative reasoning and that 15 of the exams could be passed using only imitative reasoning’ (p.348).

Table 2: Classifications of mathematics assessment questions compared to Bloom’s taxonomy and S&B’s orientations to assessment practice

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>‘factual knowledge’</td>
<td>‘reproducing bits of knowledge’</td>
</tr>
<tr>
<td>Comprehension</td>
<td>‘comprehension’</td>
<td>‘reproducing structured knowledge’</td>
</tr>
<tr>
<td>Application</td>
<td>‘routine use of procedures’</td>
<td>‘applying structured knowledge’</td>
</tr>
<tr>
<td>Analysis</td>
<td>‘information transfer’</td>
<td>‘organising subject knowledge’</td>
</tr>
<tr>
<td>Synthesis</td>
<td>‘applications in new situations’</td>
<td>‘transforming discipline knowledge’</td>
</tr>
<tr>
<td>Evaluation</td>
<td>‘justifying and interpreting’</td>
<td>‘transforming conceptions of the discipline/world’</td>
</tr>
<tr>
<td></td>
<td>‘implications, conjectures and comparisons’</td>
<td></td>
</tr>
</tbody>
</table>

Context: F2F
5.3 Summary

- Deep learning, the broader theoretical background informing studies on approaches, is discussed and shown to be related to some of the literature on the quality of learning in mathematics.
- Research into approaches is linked to the quality of student learning outcomes.
- Various studies from the approaches to teaching and assessment literature are considered as a means to investigate the nature of current FO mathematics instructors’ assessment practices.
- Two of these – Trigwell and Prosser’s (2004) approaches to teaching inventory and Samuelowicz and Bain’s (2002) orientations to assessment practice framework – are selected for use in the thesis study.
- Work on classifying mathematics questions is summarized and appears aligned with framework findings on general approaches to assessment practice.
6. The Practice of Assessment in FO Mathematics Courses

With substantive issues summarized and theoretical frameworks for analysis presented the thesis now turns to a review of the field of study: assessment practice. Assessment has been defined as ‘the process by which assessors make inferences about the learning development accomplished by students and gain information for improving teaching and learning strategies’ (Brown & Knight, 1994 as cited in Bridges et al., 1999, p.286). It has been widely assumed to have a powerful influence in directing and (Davis, Harrison, Palipana, & Ward, 2005; Houston, 2002; Marriott & Lau, 2008; Smith & Wood, 2000), as Smith and Wood (2000) contend, ‘driving’ student learning. For this reason, studying instructors’ assessment practices may be considered an important means of gaining insight into the target of instruction as well as the opportunities instructors have to develop student understanding.

However, despite such recognition, assessment practice is often overlooked and there is relatively little research on the actual practice of assessment (e.g. Baume, Yorke, & Coffey, 2004). In F2F tertiary mathematics, Burn, Appleby, and Maher (1998), in their book titled Teaching undergraduate mathematics, devote only 16 of 263 pages to assessment. Concerning the FO context, the research need is even greater.

This thesis study addresses this need by investigating FO mathematics instructors’ assessment practice. This chapter provides a summary of related research with a particular focus on tertiary-level mathematics instruction. In addition, given the connection the literature makes between FO environments and the use of CAA (e.g. Hibberd, Litton, Chambers, & Rowlett, 2004), the use of CAA is a focus. The chapter begins with a general discussion, including theoretical background, on assessment practice from which feedback practice is identified for its influence on the quality of student learning. A review of the research on feedback is then provided with a particular focus on the nature of feedback and its influence on learning. Following on this, the literature on current tertiary F2F and then FO mathematics assessment practice is reviewed. Next four aspects of assessment practice are identified for the thesis research, namely the use of discussion, a variety of assessment instruments, online quizzes and
invigilation\(^8\). With the former two practices previously discussed in chapter four, the chapter concludes by discussing the latter two practices.

6.1 Theoretical Background: Assessment Practice

At least four types of assessment have been identified – formative, summative, accountable and diagnostic (Engelbrecht & Harding, 2004) – with formative and summative assessment the focus of significant interest since Scriven (1967) initially proposed the terms. Formative assessment is generally linked with feedback practices that are concerned with closing any gap in learning (i.e. a developmental process, e.g. Sadler, 1998). Summative assessment is linked with the idea of feedout (i.e. a judgement of the product, e.g. Knight, 2002) where the gap in student learning is simply measured with no specific intention to address any learning needs.

In practice, there is a good deal of confusion concerning how these terms are used (Asghar, 2012; Harlen & James, 1997; Taras, 2008). In particular, there is a recognition that the line that separates them is considered ‘blurred’ (Brookhart, 2004; Gikandi, Morrow, & Davis, 2011; Yorke, 2003). In FO mathematics, for example, ‘projects’ may fulfil both formative and summative purposes.

In relation to their value in advancing student learning, the literature generally emphasizes the untapped potential of formative assessment practices (e.g. Nicol & Macfarlane-Dick, 2006) and the limitations of summative assessment practices (e.g. Birenbaum et al., 2006). Such emphases are consistent with claims that link formative assessment practices to higher quality learning (Harlen & James, 1997) and summative assessment practices to lower quality learning (In general, Knight, 2002; In mathematics, Boaler, 1998; Houston, 2002; Solomon, 2007). While such claims have recently been questioned (e.g. Bennett, 2011), the literature does make a causal link between the quality of feedback and the quality of learning (e.g. Hattie & Timperley, 2007). For this reason, the use of feedback will be a major focus of the thesis research and the chapter will now review the literature on feedback practice in assessment.

\(^8\) ‘Invigilation’ in the UK context is equivalent to ‘proctoring’ in the US context. Both considered as forms of human supervision.
6.2 The Practice of Feedback in Assessment

Ramprasad (1983) defines 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’ (p. 4). In this thesis, feedback generally refers to information given by the instructor to the student in the context of assessment. This includes automated CAA feedback (also referred to as feedback via computer agency or computer-generated feedback) that was originally set by the instructor. This section summarizes the literature on feedback practice by providing an overview of current research on feedback and then focusing on two aspects of feedback – timing and ‘kind’.

6.2.1 Research Overview

Both empirical research findings (e.g. Hattie & Timperley, 2007) and reflective claims (e.g. Biggs & Tang, 2007; Ramsden, 2003) attest to the power of feedback. However, wide variability in the effects of feedback has been reported with some feedback found to produce no or debilitating effects on performance (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). For example, Kluger and DiNisi (1996), conducting a meta-analysis of 607 effect sizes, found over one-third of the studies showed a negative effect on student performance. Hattie and Timperley (2007) investigated the effect of feedback on student achievement by reviewing 12 meta-analyses which included 196 studies and 6,972 effect sizes. They found an average effect size of 0.79. In their overall conclusion, they suggest that the most effective feedback is directed at the task and how it can be done more ‘more effectively’ (p.84), while the least effective feedback was related to giving rewards or punishment (e.g. praising student effort). Other findings indicate differences in efficacy based on, for example, how frequently feedback is provided (e.g. number of times feedback is given over the course of attempting one question, Hattie & Timperley, 2007), how specific it is (e.g. addressing ‘correctness of the minutiae of tasks’, Hattie & Timperley, 2007, p. 91; Goodman, Wood, & Hendrickx, 2004), the nature of the subject matter being studied (Kluger & DeNisi, 1996) and the characteristics of the student receiving the feedback (Butler & Winne, 1995).

---

9 This was approximately twice the ‘average or typical effect of schooling’ of 0.4 (p.83). Feedback was also one of their top 10 greatest influences on learning.
The following sections address some of this complexity by considering some issues highlighted in the literature. In particular, two characteristics of feedback – timing and kind – are considered for their influence on the quality of learning.

### 6.2.2 Feedback Timing

In the deep learning literature it is claimed that the provision of ‘rapid’ feedback ‘promote(s) active, deep learning’ (Entwistle, 2003, p.11). However, in the general literature, the status of the effect of feedback timing, where a typical study focuses on the effects of immediate versus delayed feedback, is unclear. Instead, findings suggest that these effects are differentiated based on how difficult an individual student finds a particular question where, for example, the greater the need for students to process the material to gain understanding, the more they may benefit from delayed feedback (Kluger and DiNisi, 1996). Still, in another meta-study, Shute (2008) concludes that the benefit of immediate feedback may be limited to lower-level ‘procedural skills’ (p.165) and providing motivation. These findings suggest a greater degree of complexity than simply stating immediate (or delayed) feedback is beneficial to student learning.

In mathematics, the research appears less clear. Simmons and Cope (1993), in one of the few empirical studies in mathematics, compared the same problems presented using a computer software package with those on paper (n=64 children). Their findings suggest that immediate feedback acts to ‘inhibit moves to a higher level of response’ (p.163) and as such ‘upsets the balance’ (p.175) by better enabling procedural, to the detriment of conceptual, knowledge and understanding. In contrast, some claim immediate CAA feedback is linked with ‘gaining a thorough understanding’ (p.56) of mathematics (Zerr, 2007) and others have found it linked with increased grade performance (Butler, Pyzdrowski, Goodykoontz, & Walker, 2008). However, it is unclear this ‘understanding’ or performance refers to the kind of procedural understanding Shute (2008) refers to. And if interpreted this way, the balance of evidence on the effect of feedback timing in mathematics is consistent with findings in the general literature which suggest immediate feedback would be beneficial to procedural learning whereas delayed feedback would be beneficial to conceptual learning.

### 6.2.3 Feedback Kind

In the deep learning literature, Biggs and Tang (2007) claim that the effectiveness of any particular instruction is directly related to the quality of feedback provided. Similarly, Trigwell,
Prosser and Waterhouse (1999) have linked instructor-provided ‘helpful feedback’ to the likelihood that students ‘report adopting a deep approach’ to learning (p.66).

Kinds of assessment feedback may be characterized in several ways. One granular representation of findings in the literature may be a taxonomy that relates both claims and findings concerning the effects of three kinds of assessment feedback on the quality of student learning. The first kind, considered the poorest quality assessment feedback, consists solely of a grade or mark. Such feedback is widely considered (e.g. Taras, 2002) and found (e.g. van der Kleij, Theo, Timmers, & Veldkamp, 2012) to be the least beneficial to student learning. For example, it does not provide any direction to help further learning (Kvale, 2007) and is possibly detrimental (Sadler, 1989). At the other end of the continuum, the kind of feedback widely considered (e.g. Entwistle, 2009) and found (e.g. Hattie & Timperley, 2007) to be most beneficial is hints and comments directed at the learning process. That is, feedback directed at the learning task (Hattie & Timperley, 2007) and the development of student understanding (e.g. Entwistle, 2009). The final intermediate kind of feedback consists of providing the correct answer or a full solution (whether computer- or instructor-generated). This feedback was found to be better than the first category (e.g. Kluger & DeNisi, 1996) but not considered to be better than the third category (Ramsden, 1988 as cited in Anderson & Garrison, 1995).

This taxonomy (see Table 3) is consistent with findings related to mathematics. For example, in statistics, CAA feedback ‘providing the correct answer was found to be superior to feedback simply saying whether the student’s answer was correct or wrong; and this in turn was found to be superior to the total absence of feedback’ (Roper, 1977, p.43). Additionally, in geometry, CAA feedback that provided hints was found to significantly improve learning, as measured by pre to post test gains, than feedback that showed answer correctness or the full solution (Singh et al., 2011). Finally, the only known meta-analysis on feedback practice in mathematics found that specific feedback on performance was found to consistently ‘enhance mathematics achievement’ (Baker, Gersten & Lee, 2002, p.67). The taxonomy in Table 3, presenting a framework for analyzing FO mathematics instructors’ assessment practices, will also be used in this thesis study.
Table 3: A proposed taxonomy of kinds of feedback in relation to the effect on learning

<table>
<thead>
<tr>
<th>Description of Feedback</th>
<th>Quality of Feedback</th>
<th>Related Basis in the General Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct/Incorrect</td>
<td>Poor</td>
<td>(Kvale, 2007; Sadler, 1989; Taras, 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(van der Kleij et al., 2012)</td>
</tr>
<tr>
<td>Full solution</td>
<td>Intermediate</td>
<td>(Ramsden, 1988 as cited in Anderson &amp; Garrison, 1995)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Kluger &amp; DeNisi, 1996; Moreno, 2004; Singh et al., 2011)</td>
</tr>
<tr>
<td>Hints or comments</td>
<td>Rich</td>
<td>(Entwistle, 2009; Kluger &amp; DeNisi, 1996; Laurillard, 2002; Merrill, Reiser, Ranney, &amp; Trafton, 1992; Ramsden, 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Hattie &amp; Timperley, 2007; Moreno, 2004)</td>
</tr>
</tbody>
</table>

Overall, while the issue of the effect of feedback timing on learning appears less settled, the limited evidence is consistent with feedback kinds, such as hints and comments, as most effective for quality learning in mathematics. R2 of the first study will investigate feedback timing and kind. Furthermore, claims and findings concerning the effects of different kinds of feedback will be used to create a framework for analysing instructors’ assessment practices.

6.3 F2F Mathematics Instructors’ Assessment Practices – Instruments, Weighting and Delivery

The following section provides a brief review of research regarding the assessment schemes\(^\text{10}\) used in F2F mathematics courses. This provides some perspective on how assessment practices may have changed as a result of adapting them to the FO instructional context.

Despite claims (Brown & Knight, 1994) and findings (Wormald, Schoeman, Somasunderam, & Penn, 2009) of their power in directing student learning, little research examines the way assessment instruments are combined (i.e. weighted) into an overall course assessment scheme in a mathematics course. For example, neither the Mathematical Association of America (MAA) book titled Assessment Practices in Undergraduate Mathematics (Gold, Keith, & Marion, 1999) nor the Higher Education Funding Council for England More Maths Grads

\(^\text{10}\) Here we define assessment schemes as the instruments used together with their associated weighting.
Project book titled *Maths at University: Reflections on experience, practice and provision* (Robinson, Challis, & Thomlinson, 2010) presents any significant research findings or focus on assessment schemes. This is surprising given claims regarding the power of assessment in driving and directing learning (Boud, 1990; Davis et al., 2005; Houston, 2002; Marriott & Lau, 2008; Smith & Wood, 2000), which is assumed to be operationalized through the use of assessment schemes (e.g. Wormald et al., 2009).

Complicating and limiting research into these schemes are problems with terminology. In particular, assessment components that are classified using the same or different terminology may actually serve different or the same purposes. For example, [www.merriam-webster.com](http://www.merriam-webster.com) defines homework as ‘an assignment given to a student to be completed outside the regular class period’ and quizzes as a ‘short oral or written test’. Homework would typically be considered a formative-style assessment instrument. The thesis defines formative-style assessment instruments as those instruments given during planned instruction and with feedback assumed to be focused on the process of learning what is being instructed. However, is a small project considered a homework assignment? Is CAA homework essentially the same as CAA quizzes if the same problems are used but the homework is graded by the problem while quizzes are graded as a whole? Another example is whether a chapter or unit test would be considered a summative-style assessment instrument. This thesis defines summative-style assessment instruments as those instruments given at the end of planned instruction and with feedback assumed to be focused on the product of learning what was instructed. However, when does a test become a quiz? In short, the role of a particular assessment instrument may not be clear and may vary from instructor to instructor.

Moreover, apart from these distinctions, the way an assessment instrument is delivered also varies. For example, summative-style assessment instruments are typically assumed to be delivered under controlled conditions (e.g. time, human supervision). However, formative-style assessment instruments may, for example, not be supervised. Additionally, assessment instruments may be paper-based or delivered via CAA. And if CAA is used, the nature and ability to change questions or feedback may be constrained by the computer platform. In short,

---

11 Use of this terminology (i.e. formative- and summative-style) appears more in keeping with actual practice where, as Bennett argues (2011), an instrument may have a ‘primary’ and a ‘secondary’ purpose.

12 Within a single course, one could argue that ‘final exams’ are the only exception. The inclusion of other ‘exams’ or ‘tests’ is considered debatable given it is unclear to what degree these exams are intended to just measure or also close the learning gap. For example, in preparation for the final exam, if retakes are available or if corrections can be submitted for extra credit.
the way a particular assessment instrument is delivered and role it plays may not be clear and may vary not only from instructor to instructor but also from platform to platform.

An emphasis on summative-style instruments is broadly recognized in F2F tertiary mathematics (Engelbrecht & Harding, 2005b; Houston, 2002; Iannone & Simpson, 2011; Ross, 1999; Solomon, 2007; Wood & Smith, 1999). First, in the UK context, Iannone and Simpson (2012b) present one of the few empirical studies investigating assessment schemes. They systematically analysed 43 BSc Mathematics degree programs (‘courses’ in UK terminology) involving 1843 mathematics courses taught in mathematics departments across England and Wales. They found a ‘system dominated by the closed book examination’ (p.13) with the grades from these exams having a median contribution of 72% towards the final degree (i.e. averaged across all courses in each program) and few departments seeing this contribution at less than 50%. Second, also in the UK context, Challis, Robinson and Thomlinson (2010) found coursework weightings varied from 15% in the first year of university to nothing in later years. Third, once more in the UK context, Green, Harrison, Mustoe and Ward (2003), in their UK-wide work on helping engineers learn mathematics, state that coursework has a ‘typical’ weighting of 30% whereas a written examination has a ‘typical’ weighting of 70%. Fourth, in the US context, Bonnice (1999) provides a sense of what instruments are used and how they may be weighted. In his ‘flexible grading’ scheme, he suggests providing students with the following choice of instruments and weighting options: self-evaluation (5%), teacher-evaluation (5%), board presentations (5 to 10%), journal (5 to 20%), projects (10 to 30%), homework (5 to 20%), hour exams (30 to 45%), final exam (15 to 25%; i.e. ‘exams’ constituting 45 to 70%). Here regional differences emerge where UK practice makes a clear distinction between two forms of assessment (i.e. ‘coursework’ and ‘examinations’) US practice refers to a variety instrument types. However, despite any such differences, both are consistent with an emphasis on summative-style assessment instruments.

---

13 Two broad categories of mathematics assessment are generally used in the UK context: examinations, which generally pertain to assessment ‘undertaken in strict formal and invigilated time-constrained conditions’ and coursework which refers to ‘all other modes of assessment’ (Bridges et al., 2002, p.36).
14 Speaking to the complexity of international research on this aspect of assessment practice, as alluded to by the former study, the UK assessment scheme is such that overall grades from different years contribute differently to the overall degree grade. For example, it is not uncommon that grades from the first year of UK university studies do not count towards the overall degree grade (cf. Iannone & Simpson, 2011). In contrast, with the US system of grade point averages, every course grade counts almost equally to the overall degree grade. This is apart from the issue of passing grades which is generally 40% in the UK and at least 50% in the US.
This study attempts to provide a measure of the quality of instructors’ overall assessment schemes. To do this, rather than relying on the use of equivalent terminology, assessment instruments are distinguished based on the kind of feedback they provide with their weighting assumed to influence how students direct their learning. That is, using the taxonomy of kinds of feedback covered earlier (see Table 3), this study considers which instruments provide which kind of feedback and combines the weighting associated with each kind of feedback. It thus provides three values considered as feedback measures of the overall quality of an individual instructors’ assessment scheme. This will be discussed further in the first study methodology.

6.4 FO General and Mathematics Instructors’ Assessment Practices

As in the F2F context, assessment in the FO context is also claimed to direct or ‘drive’ student learning (Lingefjärd & Holmquist, 2002). However, beyond this similarity, FO assessment practice is considered to be different (Allen, 2003; Gadanidis et al., 2002; Lingefjärd & Holmquist, 2002; Shuey, 2002) and more complex than F2F practice (Lingefjärd & Holmquist, 2002). Differences appear centred, for example, on how formative-style assessment instruments and feedback are emphasized. And complexity exists, for example, in how assessment practices are used to mediate for the effects of students and instructors being physically separated.

Apart from discussing these differences and complexities, this section also identifies some of the characteristic assessment approaches of general and FO mathematics assessment practice. First, a review of general FO course assessment practice is provided. Second, this is followed by a specific review of general FO course feedback practice. Third, the literature on current FO mathematics course assessment practices is considered. Fourth, two characteristic issues of current FO mathematics course assessment practice are reviewed: the use of quizzes and the issue of academic integrity. Together, this background helps frame and direct the thesis study.

6.4.1 General FO Instructors’ Assessment Practices

In general FO instruction, effective assessment practice is often linked to the use of formative-style (Anderson, 2008; Jarmon, 1999; Liang & Creasy, 2004; Walker, 2007) and a variety of assessment instruments (Gaytan & McEwen, 2007; Gikandi et al., 2011; Jarmon, 1999; Robles & Braathen, 2002; Shuey, 2002). In particular, both of these emphases are linked to higher quality learning (for formative practices, see Gikandi et al., 2011; for a variety of instruments, see Harlen & James, 1997; Ramsden, 1997). However, given the status of these claims is
uncertain, the present research sets out to investigate them in the FO mathematics course context. In the first study, R2 and R8 will investigate the relationship between FO mathematics instructors’ use of feedback and a variety of assessment instruments and their approaches to teaching and assessment. After this, R4 of the second study seeks to explain the findings related to the use feedback.

6.4.2 General FO Instructors’ Assessment Feedback Practices

The use of feedback also features prominently in several papers on FO assessment practice (Arend, 2007; Gaytan & McEwen, 2007; Liang & Creasy, 2004; Robles & Braathen, 2002; Walker, 2007). For example, in contrast to F2F instruction where the use of feedback is seen to be declining (Gibbs & Simpson, 2004), in some FO contexts such as Open Universities, it is claimed ‘students may receive fifty times as much feedback on assignments over the course of an entire degree programme as do students at conventional [i.e. F2F] universities’ (Gibbs & Simpson, 2004, p.9).

Reasons why feedback is emphasized are at least two-fold. First, as with the F2F context, good feedback practice is linked with improved quality of learning. For example, Gikandi, Davis and Morrow (2011) review ‘online formative assessment’ and conclude that online formative feedback is beneficial to higher quality learning. Second, good feedback practice is seen as a vital means of mediating for the nature of the FO course context where students are separated from each other and the instructor in both space and time (Evia, 2006; Sakshaug, 2000). Also, in the FO context, instructor feedback may be the primary contact students have with their instructor (Lingefjärd & Holmquist, 2002). And perhaps more importantly, good FO assessment feedback is considered as a means to ‘stimulate’ (Gikandi et al., 2011, p.2341) and ‘maintain’ interaction (Comeaux, 2005 as cited in Austin, 2007, p.81). As Semião (2009) argues with regards to FO mathematics courses: it serves to ‘keep students engaged with the process...of learning’ (p.439). This emphasis contrasts with general F2F assessment practice where feedback is recognized more for its role in improving than directing student learning.

15 While they do not directly define ‘formative feedback’, they do place it within the definition for formative assessment as ‘iterative processes of establishing what, how much and how well students are learning in relation to the learning goals and expected outcomes in order to inform tailored formative feedback.’ (p.2337)
Finally, regarding the role of feedback timing, Swan (2003), in her review of ‘learning effectiveness’ in FO courses, characterizes effective online feedback as ‘prompt’, ‘constructive’, ‘frequent’ and ‘clear’ and then goes on to link effective online learning to ‘ongoing assessment of student performance... [with] immediate feedback and individualized instruction’ (Swan, 2003, p.24). This also contrasts with general F2F assessment practice where immediate feedback is seen as a possible threat to higher quality learning.

In summary, in general FO courses, feedback plays a vital role. Beyond improving student learning, it is also seen to help mediate for the physical separation of the instructor and students in both space and time. R2 of the first study will investigate how current FO mathematics instructors are using feedback in their FO mathematics courses. After this, R4 of the second study seeks to explain, through interviews with a sample of US survey participants, how and why feedback is being provided.

### 6.4.3 FO Mathematics Instructors’ Assessment Practices

There is a growing body of literature on the use of formative-style assessment instruments in FO mathematics courses (e.g. for peer, see Everson & Garfield, 2008; group work or projects, Lingefjärd & Holmquist, 2002; discussion, Pomper, 2007; journal, Tsvigu, Breiteig, Persens, & Ndalichako, 2008). However, their use in FO mathematics practice appears limited. Instead FO mathematics course assessment practices mirror F2F practices where summative-style instruments prevail.

Evidence of similarities may be found in at least three studies. First, Galante (2002) analyzed 3 interviews and email survey responses from 37 (of 472 or 9% response rate) FO mathematics instructors from across the US. While her findings (see Table 4) regarding instrument usage consider a wide variety of instruments, weighting is considered for only three categories of assessment instrument: tests, ‘electronic communication’ and ‘homework and other assignments’ (p.103). Among her conclusions: ‘tests and quizzes, represented the largest portion of a student’s final grade...and (in some cases), tests were the only form of assessment’ (p.158). In addition, it was found that 16% did not use any form of invigilation. Second,

---

16 Learning effectiveness in this study is considered as a measure of how institutions FO learning is equivalent to or better than F2F learning.

17 This figure does not take into account five of the 37 participants that did not respond. It may be argued, particularly given the highly contentious nature of this issue within the mathematics community (Trenholm, 2007), that this under-represents the true number of non-invigilating instructors.
Trenholm (2007a) conducted an email survey study of 47 (of 122 or 39% response rate) FO mathematics instructors in one US state-wide VLE context. Findings suggested a process of pedagogical adaptation that departs from traditionally accepted and trusted assessment practices where tests and exams supervised by humans are a critical component of mathematics assessment. Among the main findings, the majority (64%) of courses did not use any invigilation and a wide variety of instruments (see Table 5) were used with types used and weighting largely differentiated by whether invigilation was used. That is, while those that used invigilation generally placed a similar emphasis – in terms of usage and weighting – on summative-style instruments, those not using invigilation did not. Instead, these instructors were more than 4 times as likely to use ‘projects’ and weight them significantly higher when compared to invigilated courses. He argued that instructors who did not use invigilation were replacing summative-style assessment instruments with formative-style assessment instruments. Finally he found the average number of different instruments used was four (SD=1) with little difference based on the use of invigilation. In summary, Trenholm’s findings appear consistent with the wider FO assessment practice emphasis on the use of discussion and formative-style instruments but only for courses that are not invigilated. For those courses that are invigilated, his findings are consistent with the F2F assessment practice emphasis on summative-style instruments. Third and lastly, Smith et al.’s (2008) log file analysis of over 500 US students, discussed earlier in the chapter on FO mathematics, also found that hard-pure FO courses (e.g., mathematics) use tests and question pools more than other disciplines in the FO context. In summary, with the exception of those courses where invigilation is not used, the limited findings reflect a similar emphasis on summative-style instruments found in F2F practice. R1 of the first study will re-investigate FO mathematics instructors’ course assessment schemes by asking what instruments are currently being using to assess students and how they are weighted.
### Table 4: Summary of Findings from Galante's (2002) Study

<table>
<thead>
<tr>
<th>Instrument</th>
<th>% Using</th>
<th>Avg. Weight (%)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests (Multiple Choice)</td>
<td>70</td>
<td>69</td>
<td>20.6</td>
</tr>
<tr>
<td>Tests (Problem Solving)</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests (Short-answer) or Quizzes</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic Communications</td>
<td>78</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Homework</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Group</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Individual</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolios</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research paper</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphing simulation</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5: Summary of Findings from Trenholm's (2007) Study

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Invigilated</th>
<th>Non-Invigilated</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Using</td>
<td>Avg. Weight (%)</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Final Exam</td>
<td>100</td>
<td>26.0</td>
<td>7.4</td>
</tr>
<tr>
<td>Tests</td>
<td>96</td>
<td>46.6</td>
<td>14.7</td>
</tr>
<tr>
<td>Quizzes</td>
<td>59</td>
<td>14.5</td>
<td>8.6</td>
</tr>
<tr>
<td>Discussion</td>
<td>64</td>
<td>9.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Homework</td>
<td>55</td>
<td>17.0</td>
<td>5.8</td>
</tr>
<tr>
<td>Project</td>
<td>13.6</td>
<td>18.7</td>
<td>15.0</td>
</tr>
<tr>
<td>Attendance</td>
<td>18.2</td>
<td>6.4</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Finally, there are claims, but little empirical evidence, that the use of CAA is highlighted by FO mathematics instructors. Engelbrecht and Harding (2005a), for example, refer to the use of CAA as ‘an integral part of the assessment strategy’ (p.247) for ‘internet-based’ courses. Given some claim (Ricketts & Wilks, 2001) and others have found (Varsavsky, 2004; Griffin & Gudlaugsdottir, 2006) that the use of CAA can lead to improved performance on summative-style assessment instruments, this may seem aligned with the current emphasis on summative-style assessment instruments in mathematics. As part of the research question on assessment schemes, this study also investigates the use of CAA-based assessment instruments in current FO mathematics courses.

#### 6.4.4 FO Mathematics Instructors’ Assessment Practices Identified for Investigation

There are at least four FO assessment practices that are emphasized in the literature for their effect on the quality of learning. Two practices – the use of discussion and a variety of
assessment instruments – were covered earlier and will be investigated as part of the thesis study. The other two practices to be investigated – the use of CAA-based quizzes and invigilation – are detailed in this section. In R8 of the first study, the ATI measures, the S&B framework and the feedback framework are used to investigate instructors’ use of all four of these practices.

**Use of CAA-Based Quizzes in FO Mathematics Courses**

In FO mathematics instruction, CAA-based or simply ‘online quizzes’ are considered as ‘the most basic assessment activity’ (Engelbrecht & Harding, 2005b, p.247) as well as fulfilling ‘the most fundamental role in the learning process’ (Greenberg & Williams, 2008, p.355). Greenberg and Williams (2008), for example, consider ‘practice quizzes’ as ‘where nearly all of the ‘learning’ takes place’ (p.355). These quizzes have been considered as fulfilling a primarily formative role\(^\text{18}\) (e.g. Griffin & Gudlaugsdottir, 2006) which, as such, is aligned with current general FO assessment practices.

In the literature, online quizzes are characterized by the provision of quick or immediate feedback, frequent administration (e.g. weekly) and the use of randomized questions that provide multiple attempts to answer (i.e. what has been termed an 'attempt-feedback-reattempt system', Butler et al., 2008, p.132; Griffin & Gudlaugsdottir, 2006; Lowe & Hasson, 2011; Varsavsky, 2004). Hodge (2009) describes a typical\(^\text{19}\) approach to using quizzes: ‘Individual participants took varied numbers of practice quizzes, but a trend of taking practice quizzes until 3 to 5 consecutive practice quizzes were completed with scores of 100% was observed. For some, only 3 to 4 practice quizzes were needed to achieve this level of proficiency, while some reported taking as many as 15. This core strategy was used consistently throughout the semester…Most students tracked their progress week-to-week using the online quiz and test system to monitor their progress’ (p.235).

There are competing claims and findings concerning the potential of online quizzes to effect learning. These may first be understood by considering some of the wider CAA literature in mathematics. Some consider the use of CAA as beneficial to mathematics instruction (e.g. Pitcher et al., 2002), for example leading to improved performance (e.g. Varsavsky, 2004). In

\(^{18}\) Quizzes present an example of blurring between formative and summative purposes (e.g. Brookhart, 2004). As short tests that provide a measure of performance, they may be considered as a summative-style instrument (e.g. Knight, 2002). However, given the emphasis on feedback and attempts to address the learning gap (e.g. Ramaprasad, 1983), they can also be considered as a formative-style instrument.

\(^{19}\) See also (Greenberg & Williams, 2008)
contrast, others have found that it can only address the two lowest levels of Bloom’s (1956) taxonomy (i.e. knowledge and comprehension) and it is doubtful that it can ‘soundly test the remaining four levels’ (Paterson, 2002, p.302).

The literature concerning the use of online quizzes in mathematics mirrors and further extends these claims and findings, particularly regarding the combined provision of immediate feedback with multiple attempts. A recent large scale study is illustrative. Angus and Watson (2009) analyzed administrative (e.g. prior student aptitude) and assessment data of students (n=397 students for the first sample and 1239 for the second) using online quizzes in a business mathematics course at one Australian university. Feedback ‘in the form of the correct final answer was immediately given to students on completion of their whole attempt’ (p.261). Using regression analysis they found a ‘higher exposure to the online instrument [i.e. online quizzes] robustly leads to higher student learning’ (p.255).

This study raises at least a couple of issues: First, given ‘higher student learning’ was measured by ‘attainment of a higher final examination mark’ (p.262), it is unclear what quality of learning is being measured. That is, it is possible these gains are only limited to developing lower-level procedural understanding, as may be expected with the use of immediate feedback (e.g. Shute, 2008). Second, though students in this study were limited to two quiz attempts (counting only the better of the two grades), similar learning processes with ‘repetitive examples’ have been linked to the potential reinforcement of ‘incorrect interpretations’ (Dubinsky, 1991, p.28) or what Sangwin (2010) refers to as an ‘automatic strategy with no underlying mathematical understanding’ (p.243). In fact, as Simmons and Cope (1993) found, these environments can encourage ‘trial-and-error strategies’ that keep students focused on ‘low-level’ thinking (p.175).

In short, it remains unclear what kind of learning online quizzes support. R1 and R8 of the first study will investigate how quizzes, in general, are being used in FO mathematics courses. R2 of the second study then seeks to explain, through interviews with a sample of US survey participants, how and why quizzes are being used.

**Use of Invigilation in FO Mathematics Courses**

Several papers have raised concerns regarding test security in online mathematics (Allen, 2003; Mascuilli, 2000; Semião, 2009; Trenholm, 2007a). In FO courses it has been noted that instructors might feel under pressure to uphold the ‘anytime, anywhere’ creed of FO instruction (Trenholm, 2007b), particularly where students may demand that flexibility (Summers et al., 2005). However, students might feel motivated to cheat given the anonymity afforded by the
FO context and the relative ease of cheating in mathematics (Smith, Torres-Ayala et al., 2008; Trenholm, 2007a; Trenholm, 2007b). For example, while cheating may not be an issue in soft disciplines where assessment instruments already naturally tend to be unsupervised (e.g. term paper), this is considered an issue in mathematics given the typical emphasis on the controlled administration of summative-style assessment instruments (e.g. exam). Additionally, in contrast with soft disciplines where student responses tend to be subjective and writing-based (i.e. providing an ‘intellectual fingerprint’ of students work), cheating is seen to be easier to commit and harder to detect in mathematics where responses tend to be objective and numerically or symbolically-based (Trenholm, 2007b). Together these issues create tension between the convenience and flexibility demands of students and institutional needs to uphold standards of quality (Trenholm, 2007b). Overall the use of invigilation is less an issue of supporting quality learning and more an issue of the validity of the process of how that learning is being measured.

Within the mathematics community the issue of non-invigilation is a contentious one (Trenholm, 2007a) with some arguing for (Flesch & Ostler, 2011) and against (Yates & Beaudrie, 2009) the need for test supervision. First, Flesch and Ostler’s (2011) compared two groups of US community college students taking an intermediate algebra course. Students were randomly assigned to either invigilated (n=30) or non-invigilated (n=32) test formats. In their analysis they found that students in the non-invigilated condition did significantly better than the students in the invigilated condition and concluded that non-invigilation leads to grade inflation. Second, Yates and Beaudries (2009) study, analysing 850 FO mathematics course grades (with 406 students tested in a supervised environment, and the remaining 444 in an entirely online unsupervised setting), found that there was no significant difference in the two groups concluding that unsupervised testing is a ‘reasonable approach’ (p.68). In keeping with methodological problems in much of the wider FO research, the latter study prompted a response article (Englander, Fask, & Wang, 2011), outlining five problem areas which effectively nullified their findings. As reflected in these studies, the balance of research appears to suggest a need for invigilation. R3 of the first study will investigate the use of invigilation. It will also expand on previous studies by investigating which assessment instruments are invigilated. Following this, R3 of the second study then seeks to explain, through interviews with a sample of US survey participants, how and why participants are choosing to use invigilation.
6.5 Summary

- Assessment practice is identified for its claimed potency in directing student learning.
- The theoretical background on assessment practice reveals some confusion concerning how terms are operationalized. However, feedback practice is identified for its influence on the quality of student learning with the timing, kind and computer-generation of feedback selected for further study.
- In mathematics, the status of the effects of feedback timing on learning appears unclear whereas the status of the kind of feedback appears more certain. Findings and claims from the literature suggest a taxonomy that relates the influence of three kinds of feedback on the quality of student learning. This presents a further framework which will be used to analyze the quality of instructors’ assessment practices.
- With the possible exception of FO courses that do not use invigilation, summative-style assessment instruments are emphasized in current tertiary F2F and FO mathematics assessment practice.
- Four characteristic FO assessment practices are identified from the literature for their relationship to student learning and for study in the thesis: use of discussion, feedback, quizzes and invigilation.

6.6 Literature Review Summary

In summary, assessment practice is claimed to have a powerful influence on student learning and the FO mathematics instructional context identified as a needy area of research. This need is considered particularly great for at least three previously discussed reasons. First, most of the current FO instructional research fails to consider the disciplinary context. Second, several characteristics of current mathematics teaching and assessment practice do not fit with those practices currently emphasized in the FO instructional context. Third, most ‘efficacy’ research uses quantitative learning outcome measures where the quality of that learning is unclear. To address this research gap three study frameworks – each addressing some aspect relating to the quality of teaching and assessment practice – are used to investigate current FO mathematics instructors’ assessment practices.
7. Thesis Research Design

7.1 Introduction

The main overarching research question – what is the nature of current FO mathematics courses? – is addressed through a primary focus on FO instructors’ assessment practices and by conducting a mixed methods study. The first study, using quantitative research methods, asks what specific assessment practices are used in these courses and whether some of these practices are related to measures of instructors’ approaches to teaching and assessment. Then, directed by the first study, the second study uses a qualitative research approach to ask how and why some assessment practices are used, how individual participants are approaching their assessment practice and how these results reflect on the quality of learning in FO courses. Overall 13 research questions are asked: eight for the first study and five for the second.

The following chapter discusses the overarching mixed methods research approach used in the thesis. First, a theoretical justification for a mixed methods study is offered. Following this the choice of research design is discussed in relation to the thesis research.

7.2 Mixed Methods Research: Theoretical Background

According to Creswell (2007), the central premise of mixed methods research is that ‘the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone’ (p.5). This argument appears to be taking root in the research community as evidenced by reports of a steadily increasing number of studies integrating quantitative and qualitative research (Creswell, 2007; Bryman, 2006).

Creswell (2007) traces the historical development of mixed methods research as far back as Campbell and Fiske’s (1959) seminal paper, Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix, which advocated for a process of validation through the use of multiple methods of measuring psychological traits. As others followed (Sieber, 1973; Jick, 1979; Cook and Reichardt, 1979; all as cited in Creswell, 2007) a debate arose that discussed the viability of combining what to many were disparate research approaches (e.g. Rossman and Wilson, 1985).

At the heart of this debate is what Johnson and Onwuegbuzie (2004) refer to as quantitative and qualitative purists that each ‘view their paradigms as the ideal for research’ (p.14) and advocate
for Howe’s (1988) ‘incompatibility thesis’ which argued that these approaches cannot be mixed. Quantitative purists, according to Johnson and Onwuegbuzie (2004) believe ‘that social observations should be treated as entities in much the same way that physical scientists treat physical phenomena’, further contending ‘that the observer is separate from the entities that are subject to observation...[and] that social science inquiry should be objective’ (p.14). Consistent with positivism, quantitative research is characterized by, for example, testing theories and hypotheses through the use of statistical analysis. In contrast, qualitative purists reject the detached observer status of quantitative research and ‘contend that multiple-constructed realities abound, that time- and context-free generalizations are neither desirable nor possible, that research is value-bound, that it is impossible to differentiate fully causes and effects, that logic flows from specific to general (e.g., explanations are generated inductively from the data), and that knower and known cannot be separated because the subjective knower is the only source of reality (Guba, 1990)” (p.14). In contrast with quantitative research, qualitative research is characterized by, for example, theory and hypothesis generation through the use of qualitative methods of analysis.

This debate appears to have subsided with some now advocating for mixed methods research as ‘a research paradigm whose time has come’ (Johnson and Onwuegbuzie, 2004, p.14) or a ‘third methodological movement’ (Creswell, 2007). The growing acceptance of and advocacy for mixed methods research appears to have come about for at least two reasons. First, qualitative and quantitative researchers have been able to settle many of their disagreements. For example, researchers from both sides have come to a better understanding of the subjective nature of their research (e.g. ‘value-ladenness of inquiry’; Johnson and Onwuegbuzie, 2004). Second, a philosophical argument, pragmatism, has been advanced as a means of blending both paradigms into a productive research approach, thus further bridging differences. In practice this means that researchers mix quantitative and qualitative research approaches according to the research context and which method will best help answer the research question(s). While this debate is expected to continue, the historical development of mixed methods research appears to have moved into a period of recognition as a valid research approach (Creswell, 2007).

In advocating mixed methods research as a practicable research approach, many value it for combining the strengths and counteracting the separate weaknesses of quantitative and qualitative research approaches (e.g. Harris and Brown, 2010). Creswell states ‘[t]he argument goes that quantitative research is weak in understanding the context or setting in which people talk. Also, the voices of participants are not directly heard in quantitative research. Further,
quantitative researchers are in the background, and their own personal biases and interpretations are seldom discussed. Qualitative research makes up for these weaknesses. On the other hand, qualitative research is seen as deficient because of the personal interpretations made by the researcher, the ensuing bias created by this, and the difficulty in generalizing findings to a large group because of the limited number of participants studied. Quantitative research, it is argued, does not have these weaknesses’ (p.9). In the end, when the research design is directed by a specific question(s), it is argued that many of these questions are ‘best and most fully answered’ by using a combination of these approaches (Johnson and Onwuegbuzie, 2004, p.18).

This is not to say that mixed methods research does not have issues that need to be addressed. For example, the pragmatic choices of mixing methods to ‘best help’ answer research questions need to be made clear. Researchers need to learn how to use and mix multiple research methods and this can mean the research process is more expensive and time consuming (Johnson and Onwuegbuzie, 2004). Nevertheless, such challenges appear to be outweighed by the identified strengths of combining approaches in service of gaining a greater understanding of research problems.

Finally, several mixed-method designs have been suggested. Leech and Onwuegbuzie (2009), for example, suggest eight basic designs that differ according to whether a study is ‘partially’ or ‘fully’ uses mixed methods, how a paradigm is emphasized and when it is used. For example, the majority of mixed methods studies employ a combination of a quantitative analysis with data collected using a structured interview or questionnaire and a qualitative analysis with data collected using either a semi-structured or unstructured interview (Bryman, 2006). And with this basic mix studies can be conducted concurrently or sequentially.

7.3 Choice of Research Design

For the present thesis, the journey to adopt a mixed methods approach started with an initial qualitative pilot study (detailed later in the next chapter). Based on the outcome of this study, an initial quantitative research approach was decided upon given the nature of the data focused upon, at least in the early stage of the thesis, was largely quantitative. This included assessment weightings and, with early identification of potential frameworks for use in the thesis, a scale measure of approaches to teaching. In the end, only after completing the first quantitative study did a second qualitative study, and thus use of a mixed methods approach, emerge as an appropriate follow up to the first study. The choice of this approach was considered
advantageous for the primary reason that the second study may help to explain what was explored as well as, for example, any statistical associations discovered in the first study. In particular, because it was not possible to obtain a non-random sample, it was further considered advantageous, as an aid to generalizability, to use the quantitative findings to direct the selection of participants for the qualitative study. As a whole, consistent with assertions in the mixed methods research literature, the combination of a first quantitative study followed by a second qualitative study was considered an effective means of providing a greater understanding of participants’ experiences in designing their FO as compared to their F2F courses, and thus the nature of current FO mathematics courses.
8. Study I Methods and Methodology

8.1 Introduction

This first exploratory study is intended to provide direction for further research in general and the second study in particular. The following section details the seven step procedure followed for the first study. First, an initial qualitative pilot study was performed to investigate a potential methodological approach. Second, based on that pilot, a decision was made to undertake a study using an online survey and two theoretical frameworks were selected. Third, a pilot survey was created and administered to a small group of potential participants. Fourth, based on this pilot survey, changes were made to the survey instrument and a plan was created for administering the full survey. Fifth, the survey was administered to selected potential participants. Sixth, analysis was conducted on the survey data using the two frameworks and a third which was added later. Seventh, these results were reported and a second study was proposed to further answer the research questions. This methodology section provides an in depth explanation of the first six steps. It is structured in the following order: initial qualitative pilot, theoretical frameworks used in the study, pilot surveys and survey construction, full survey data collection procedure, analysis used for the survey study and concluding with a discussion on validity and reliability issues.

8.2 Initial Qualitative Pilot Study

An initial qualitative pilot study was conducted with the purpose of directing the development of the thesis study in answering the research questions. This pilot involved semi-structured interviews conducted in June 2010 with two UK university mathematics lecturers: one who taught statistics and one with significant experience using CAA. Participants were asked general open-ended questions – about how they conduct assessment in their course(s) – and more specific questions – concerning the types of assessment instruments and questions used. A concluding question asked how they defined ‘success’ in terms of an individual student’s learning outcome. Ample time was provided to share any comments or reflections. Based on this initial pilot study, where assessment weighting emerged as a variable of interest, and coupled with identification of potential frameworks, involving scale measures of approaches to teaching, it was decided to pursue a quantitative study using an online survey. Such an approach
was considered an efficient means of exploring current assessment practices of a sizable sample population of FO mathematics instructors.

8.3  Theoretical Background

As the research questions were largely focused on how FO mathematics instructors approach their assessment practice, and more specifically the quality of learning sought through those approaches, three potential frameworks for analysis were identified. The first framework was derived from a single study that investigated instructors’ ‘orientations to assessment practice’ (Samuelowicz & Bain, 2002). The second framework provided an established psychometric instrument created to measure instructors’ approaches to teaching. The third and final framework was added after the survey was closed. It was created based on both claims and findings related to the effect of assessment feedback on the quality of student learning – where the kind of feedback (e.g. grade only, full solution, hints or comments) was the basis for the creation of this framework. The following section reviews the former two and presents the last framework.

8.3.1  Use of the S&B Study Findings as a Framework for Measuring Instructors’ Approaches to Assessment

As discussed in the literature review, the main finding from the study conducted by Samuelowicz and Bain (2002) was presented as a matrix entitled ‘six assessment belief dimensions and their constituent beliefs’ ordered according to the emphasis on knowledge reproduction (hereafter KR) versus knowledge construction/transformation (hereafter KC; p.182-183; see Table 1). This matrix is used to create the S&B questions for the online survey (see Appendix B) and then used to analyse overall instructors’ approaches to assessment practice. It is also used to create a novel summated scale measure of these approaches which will be used for exploratory statistical analysis.

8.3.2  Use of the Approaches to Teaching Inventory (ATI)

As discussed in the literature review, the ATI is an established psychometric instrument used to measure how instructors are approaching their teaching practice. The 16 question version of the ATI (see Appendix A) was used largely as Trigwell and Prosser (2004) instruct. Apart from
validating the S&B summated scale measure created for this study, the ATI was used to help understand how approaches to teaching may be related to specific assessment practices.

8.3.3 A Framework for Distinguishing Assessment Instruments Based on the Kind of Assessment Feedback Provided

As discussed in the literature review, there is an emphasis in the current literature on feedback as a critical characteristic that distinguishes assessment instrumentation (e.g. Taras, 2005). Research considers feedback about correctness (e.g. ‘correct/incorrect) as ‘poor’, feedback that provides hints or comments directed at understanding as ‘rich’ and feedback that provides a full solution as an ‘intermediate’ form. These distinctions were also used for analyzing the nature of instructors’ assessment practice.

8.4 Survey Build and Pilot

This section details the process by which the survey instrument was developed. It covers the iterative review process that produced an initial survey instrument to the administration of this survey in a pilot format, and lastly to the development of the final survey instrument that was used in the study.

8.4.1 Survey Instrument Build

The pilot study helped shape the initial development of the final survey questionnaire (Fink, 2003). Two issues, in particular, emerged. First, the order of presentation of questions was considered critical (Andrews, Nonnecke, & Preece, 2003). This led to the use of the S&B questions prior to questions about assessment specifics. It was thought that this would help to capture participants’ actual views and avoid questions that may act as psychological prompts that would potentially ‘set participants up’ for answering the S&B questions. Secondly, the wording of questions became an early concern (Krosnick & Presser, 2010). This, for example, became an issue of practical significance given participants were to originate from both sides of the Atlantic. The ATI was then considered for use alongside the S&B framework. The final survey was structured to fit a natural progression from questions about demographics – to help participants contextualize the survey focus – followed by one question setting the specific course context – to help contextualize all remaining questions – followed by the six S&B questions, then questions on assessment specifics on grading and feedback and finishing with the 16-
question ATI. In its final format, the survey was broken into 4 sections: Demographic information, S&B questions, assessment specifics and then ATI questions (see Appendix C for final version). The latter three were all directed at a single course context.

The survey questions and instructions were then scrutinized leading to further changes. First, wording was further tested from a cross-cultural perspective. Two UK university lecturers, originally from outside the UK, were asked to review the survey and provide any feedback. This led to, for example, consideration of how ‘courses’ vs. ‘modules’ are interpreted in a US vs. UK context. Second, questions were mapped onto the research questions to help in the removal of any superfluous questions and ensure questions adequately addressed the research questions. This led to the addition of further questions regarding how participant assessment practice may be regulated by their department or institution and the removal of questions concerning participant history of teaching hybrid/blended mathematics courses. Finally, a pilot survey was conducted.

8.4.2 Procedure for Pilot

The pilot survey was launched December 2010 using Bristol Online Surveys (BOS; http://www.survey.bris.ac.uk/). It was administered to two US participants who taught both F2F and FO courses and three UK participants who taught only F2F courses. Pilot participants were asked to record the length of time it took to complete the survey and provide any feedback regarding possible areas of confusion or any further advice/thoughts. Based on this feedback, the survey underwent additional changes as detailed below.

8.4.3 Pilot Survey Feedback and Changes Made to Final Survey

General feedback concerned the length of the survey, wording of some of the questions and difficulties reading and answering the ATI and S&B sections of the survey. This section details this feedback with action(s) taken.

1. Survey participants reported completion times between 18 – 30 minutes. This was considered to be too long.

Action(s) taken: Additional efforts were made to reduce the length of the survey. For example, to the demographic section, the question asking for the ‘typical number of students’ in each of the participants courses was omitted. Also, perhaps more
substantially, this was the reason the 16 question version of the ATI (vs. the originally selected 22 question version) was adopted.

2. Problems with answering S&B questions.
   Action(s) taken: Because participants found it difficult to select only one of the responses, extra instructions were added that stated: ‘If you feel none of the options are satisfactory, please pick the closest possible and feel free to comment in the space provided’.

3. Problems with ATI wording and readability.
   Action(s) taken: First, because the wording for questions 6 did not seem to fit the FO format, it was changed from ‘I set aside some teaching time...’ to ‘I plan my teaching...’ Second, because the length of the ATI statements and the appearance of the table made it difficult to keep track of the questions that were being responded to and the associated likert-style responses, the table was changed to try and make the ATI easier to follow and answer.

4. Problems with instrument weighting totals not adding up to 100%.
   Action(s) taken: Because at least one participant provided instrument weightings that did not total up to 100%, additional instructions were added that specified the ‘weighting total must add up to 100%’.

In its final form the survey consisted of 20 questions (see Appendix C) in four sections. The first section focused on participant demographics (9 questions). The second section focused on the S&B approaches to assessment (6 questions). The third section focused on assessment specifics (presented as 3 questions including 1 question for comments). The fourth section consisted of the 16 question ATI (presented as 2 questions including 1 question for comments). Following completion of the survey an opportunity was provided for participants to volunteer for a follow-up interview.

8.5 Full Survey

This section discusses ethical issues and describes the procedure used to administer the survey, the participants that were emailed the survey and the response that it received.
8.5.1 Ethical Issues

As part of the study, a number of steps were undertaken to ensure that the research was conducted in an ethical fashion. First, university ethical guidelines for research were followed and an Ethical Clearance Checklist was completed and submitted to the department. Second, in relation to this checklist, on both the email invitation and in the survey introduction, participants were fully informed of the purpose of the study and how the collected data was to be used. This included notification that the collected data would be held securely and used anonymously and confidentially.

8.5.2 Procedure

The full survey was launched using Bristol Online Surveys (BOS). This was done in stages beginning March 17th, 2011 with all surveys closed by May 31st, 2011. The selection of participants was largely based on this researcher’s prior contacts and will be detailed in the next section. The roll out of the survey was accomplished using six separate survey websites corresponding to the six groups of participants selected for the survey. This was done for at least two reasons: First, if any issues with the survey failed to be addressed in prior stages, this would provide an additional opportunity to change the survey. Second, more importantly, this was thought to provide an opportunity to track the origin of the responses (albeit imprecise, given the potential for snowball effects across participant groups). Using the six website addresses, emails were sent out to potential participants.

These emails were generally sent in two stages. In the first stage, an initial email with survey link was sent out. This was generally followed, about three weeks later, by a second reminder email. The timing of these emails was done in consideration of the academic cycles of potential participants. For example, emails were sent out to potential participants in the State University of New York at the end of March, given this was considered as a slow time in their academic cycle. The hope was that emails sent at such times were more likely to receive a response.

Emails were adapted according to the context of this researcher’s personal contact with the group contact(s). For example, an email sent to one group of contacts, related to this researcher’s prior editorial work, began by addressing the nature of the prior relationship. In so doing, an attempt was made to capitalize on working relationships rather than simply make ‘cold calls’. Additionally, where large contact lists were being used, emails were sent out to two or
three participants at a time to avoid triggering institutional electronic spam filters (Yetter & Capaccioli, 2010 as cited in Loong, 2010).

8.5.3 Participants

The online survey was administered to instructors at Western tertiary institutions who taught at least one FO mathematics course in one of three types of institutions of higher education - traditional ‘bricks and mortar’ bachelor granting (e.g. university, 4 year college), traditional ‘bricks and mortar’ non-bachelor granting (e.g. North American community college) and open/online university/college. An additional ‘other (please specify)’ category was offered if participants did not consider these categories as representing the type of institution they taught in. After the closing date of May 31st, 2011 a total of 86 surveys were started with 70 complete (see Table 6).

As shown in Table 7, about half of these participants originated from traditional ‘bricks and mortar’ non-bachelor granting institutions, almost half originated from traditional ‘bricks and mortar’ bachelor granting institutions, one tenth originated from open/online university/college and a small number were either listed as ‘other’ or were not specified. Of particular importance, most participants teach in the US higher education context and, given all participants teaching in traditional ‘bricks and mortar’ non-bachelor granting institutions come from the US, it is assumed these are US community college instructors. The second study details the background context of US higher education (see section 11.1).

While the selection of participants was largely based on this researcher’s prior contacts, considerable effort was undertaken to make contact with other potential participants from developed Western countries (i.e. Europe, US, Canada and Australia). However, as previously discussed, given the nature of online surveying, it was impossible to definitively conclude the regional representativeness of this sample. What now follows is a description of each of the six groups that became the focus of the data collection.
The survey link with the most responses was directed at FO mathematics instructors within the State University of New York (SUNY). This researcher was, for several years, a FO mathematics instructor within SUNY, involved in attending related FO workshops, presenting at related conferences, as well as conducting and publishing research based on data collected from this group (Trenholm, 2007a; Trenholm, 2007b). As such, this researcher had significant understanding of this group’s demographic make-up and had developed a contact list of FO mathematics instructors. This list comprised of approximately 180 FO mathematics instructors from both traditional ‘bricks and mortar’ bachelor granting (e.g. university, 4 year college) and non-bachelor granting (e.g. North American community college) institutions. Based on this list an initial email (see Appendix D) was sent around the beginning of April 2011 with a follow-up email reminder sent approximately three weeks later. At the close of the survey, for this link, 42 online surveys were started, with a total of 37 complete.

### Table 6: Summary of online survey participants’ responses

<table>
<thead>
<tr>
<th>Country/Group</th>
<th>Survey Link</th>
<th>Initial Email (approx.)</th>
<th>Complete</th>
<th>Incomplete</th>
</tr>
</thead>
<tbody>
<tr>
<td>US (SUNY)</td>
<td><a href="https://www.survey.lboro.ac.uk">https://www.survey.lboro.ac.uk</a> + fullyonlinemath</td>
<td>2-Apr</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>US (TN/RUME)</td>
<td>fullyasynchmath</td>
<td>25-Mar</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>IGI</td>
<td>fullyonlinemathelearning</td>
<td>17-Mar</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Iberia/21st Century Project</td>
<td>mathelearning</td>
<td>7-Apr</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Australia</td>
<td>fullyonlinemathematics</td>
<td>30-Mar</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UK</td>
<td>fullyonlinemaths</td>
<td>12-Apr</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td>70</td>
<td>16</td>
</tr>
</tbody>
</table>

### Table 7: Summary of survey participants’ institutions by country/group

<table>
<thead>
<tr>
<th>Country/Group</th>
<th>‘2 yr’</th>
<th>‘4 yr’</th>
<th>‘Online’</th>
<th>Other</th>
<th>Unclear from response</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>US (SUNY)</td>
<td>25</td>
<td>7</td>
<td>0</td>
<td>3*</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td>US (TN/RUME)</td>
<td>6</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>IGI</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>1**</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Iberia/21st Century Project</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Australia</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>UK</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>31</td>
<td>26</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>70</td>
</tr>
</tbody>
</table>

*Specified by participants as a combination of 2 and 4 yr institution. This will be further discussed in the second study results.

**Specified as ‘Institute of Technology Ireland’
The next survey link was directed at two groups of contacts. The first of these two groups were FO mathematics instructors that taught within the Tennessee (TN) Board of Regents system (comprised of colleges, universities and technology centres). The email for this group was sent to a single research contact that this researcher met at a FO mathematics instruction conference in Tennessee and identified as a leading researcher within the system. The second of these two groups were members of the listserv for the Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education (SIGMAA on RUME). Because RUME was identified as the US research body focused on undergraduate mathematics and because one of this researcher’s supervisors had personal contact and was personally known by RUME, this was considered to be another important avenue for finding potential participants. Here, personal contact was first made with the individual moderating the listserv (Dr. Eric Hsu of San Francisco State University). Following this, a request for participants was circulated on the listserv. In both of these groups, contact was made with a single individual who was asked to circulate the request. The number who actually received emails inviting participation was not known. At the close of the survey, for this link, 21 online surveys were started, with a total of 16 complete.

The following survey link (IGI) was directed at a group of contacts established through this researcher’s work co-editing a book on mathematics e-learning (Juan, Huertas, Trenholm, & Steegmann, 2011). Emails were sent to each of approximately 30 research contacts who were authors of papers accepted for publication. These were researchers involved in all aspects of e-learning (i.e. as broadly defined in the literature review) and not necessarily involved in FO mathematics instruction – as evidenced by some of the email replies to this researcher. At the close of the survey, for this link, 11 online surveys were started, with a total of 8 complete.

The final three links (Iberia/21st Century Project, Australia, UK) were sent to a combination of direct (Iberia and Australia) and indirect (UK) personal contacts with several ‘cold-call’ attempts attempting to make direct contact (e.g. 21st Century Project). Iberian contacts were sought through a single individual who was a prior research collaborator. Australian contacts were attempted through two academics with prior involvement in FO mathematics research – one personally known to this researcher, the other known only through email exchanges in preparation for the Australian survey launch. Apart from these direct contacts, an additional seven Australian universities were identified, by one of these Australian academics, as significant providers of FO instruction in Australia. These universities were contacted largely via mathematics department heads and academics known to be involved in FO mathematics
instruction. Finally in the UK, considerable effort was undertaken to find UK participants both at the Open University and through five universities identified as FO course providers. The number who actually received emails inviting participation was not known. In total, at the close of the survey, for these three links, a total of 13 online surveys were started, with a total of 9 completed (Iberia/21st Century Project: 9 and 6, Australia: 3 and 2, UK: 1 and 1).

As shown, there were very few responses, relative to the effort made, for some of the links (e.g. Australia and the UK, in particular). This was thought to be attributable, for example, to the onerous time requirement (i.e. 15-20 minute completion time) or the sensitive nature of the subject matter (i.e. assessment practice in an emerging modality that is being heavily scrutinized).

In the end 70 participants completed the online survey with four participants showing some significant inconsistencies and/or missing data in the assessment specifics section of the survey. As a result these four participants were excluded from the analyses involving those responses. Data for 70 participants was available for analysis involving all 6 S&B questions (no. 10-15) and all 16 ATI questions (no. 19a-19p). However, for any analysis involving assessment specifics, data was available for only 66 participants. It should be noted that due to the design of the BOS system, participants were not forced to answer every question. This limitation will be discussed later in the methodology.

8.6 Data Analysis

The results for the first study are presented in three parts. The first part focuses on providing descriptive statistics concerning participant demographics. The second part covers the first four research questions of the first study and focuses on providing descriptive statistics concerning assessment specifics as well as comparisons of sample population groups and correlational analyses. The third part covers the last four research questions of the first study and largely focuses on comparisons of sample population groups and correlational analyses. Throughout the analyses, normality tests are run in order to determine which approach is necessary.

After detailing the descriptive statistics for the participant demographics, the bulk of this section concerns itself with explaining how the analyses were conducted for each of the research questions (denoted by ‘Study I - R1’ for the first Study I research question, ‘Study I - R2’ for the second...). Significantly, the operationalization of the proposed feedback framework and summated S&B scale measure, used in most of the analyses, are detailed. Overall, the analysis
uses a total of six, what will be termed, *study measures* consisting of the three *feedback measures* and three *approach measures*.

### 8.6.1 Participant Demographics

Initial descriptive statistics were provided for all 70 participants corresponding to questions one to six, along with question nine. This corresponded to the following demographic characteristics: role in academia, role status, institution, years of experience teaching F2F, years of experience teaching FO, FO courses taught in the last three years, survey course focus, and survey course focus level. Analysis for questions seven and eight on the survey, concerning professional development experience, is covered later in this section when the fourth research question is addressed.

Descriptive statistics provided the percentage of participants corresponding to each of the characteristics in each of the eight demographic categories. For example, the second demographic category concerning ‘status’, had three categories: full-time staff/faculty, part-time staff/faculty or neither. Percentage representation was calculated based on a total of 70 participants and rounded to the nearest whole number.

However, before descriptive statistics were calculated, the data was reviewed for problems and, where possible, fixed (see Appendix E). In addition, several responses were modified according to comments provided by participants. First, for the ‘role in academia’, one participant placed him/herself in the ‘other (specify)’ characteristic when his/her comment (‘80% teaching, 20% research’) clearly placed him/her in the ‘mostly teaching’ characteristic. This participant was reclassified under ‘mostly teaching’. Second, in the ‘institution’ category, two participants placed themselves in the ‘other (specify)’ characteristic and commented that they originated from an institution that offered both two and four-year degrees (i.e. the first two categories: offering both bachelor and non-bachelor degrees). These participants were re-classified under the first category (i.e. ‘traditional ‘bricks and mortar’ bachelor granting’).

### 8.6.2 Assessment Specifics

**Study I - R1**: What instruments are FO mathematics instructors currently using to assess their students? How are these weighted?
Assessment specifics examined when a type of assessment instrument was used and, if used, how it was emphasized as shown by its weighting and the number of separate assignments or tests associated with it over the duration of the course frequency. That is, for each type of instrument, descriptive statistics provided information on the total number of participants using the instrument and the mean and standard deviation of the weighting assigned to the instrument. Apart from these descriptives, analysis focused upon the variability as measured by the standard deviation of the weighting. Here, two assumptions were made: First, all participants generally had the same understanding regarding the definition, purpose and use of the various assessment instruments identified in the study (see 6.3 for further discussion). Second, that variability reflected the degree of ‘consensus’ participants share about how an instrument was used.

8.6.3 Use of Feedback

Study I - R2: How are instructors using feedback in their FO mathematics courses?

This section details how the feedback framework was operationalized followed by a summary of how the feedback data was analysed.

Operationalizing the Framework for Feedback

As discussed, based on findings and claims in the literature, three kinds of feedback were identified for their influence on the quality of learning. This information was then used to construct a novel scoring system to help analyze the quality of assessment instruments with respect to their potential influence on student learning. Using this framework and the survey data on feedback, each assessment instrument was classified according to its associated kind of feedback. That is, any assessment instrument providing only a grade as feedback was considered to be giving type 0 feedback and given a feedback score of zero. Those providing any or no type 0 feedback and the answer or full solution as feedback were considered type 1 and given a score of one. Those providing any or no type 0 and/or type 1 feedback with hints or comments were considered type 2 and given a score of two (see Table 8). A single participant may have acknowledged using three different assessment instruments and having provided specific kinds of feedback with each of these three instruments. They may have given the same kind of feedback for each instrument (e.g. all type 1) or they may have provided different feedback for each instrument.
Table 8: A framework, based on claims and findings discussed in the literature review, for distinguishing assessment instrumentation according to kind of feedback provided.

<table>
<thead>
<tr>
<th>Operationalization of the Framework for Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Feedback</td>
</tr>
<tr>
<td>Poor</td>
</tr>
<tr>
<td>Rich</td>
</tr>
<tr>
<td>Kind of Feedback as Described by the Survey Instrument</td>
</tr>
<tr>
<td>Grade/Mark</td>
</tr>
<tr>
<td>Correct answer provided</td>
</tr>
<tr>
<td>Computer-generated full solution/ Lecturer-generated full solution</td>
</tr>
<tr>
<td>Computer-generated hints or comments</td>
</tr>
<tr>
<td>Lecturer-generated hints or comments</td>
</tr>
<tr>
<td>Hints or comments challenge student understanding</td>
</tr>
</tbody>
</table>

This information was then used to calculate an average feedback score for each type of assessment instrument and three feedback measures for each participant. The average feedback scores for each individual kind of assessment instrument were calculated using the associated feedback scores (with weighting not considered). For example, if 28 participants used quizzes, there would be 28 feedback scores associated with the type of quiz feedback used by each participant. The average of these scores was what constituted the average feedback score for quizzes. These average scores ranged from 0 to 2, with those closer to 0 considered to provide ‘poorer feedback’ potentially leading to poorer quality learning and those closer to 2 considered to provide ‘richer feedback’ potentially leading to better quality learning. Finally, the three feedback measures were calculated for each participant as the sum total of assessment weightings associated with each feedback type. This provided a measure of how an instructor emphasized the different kinds of feedback relative to their overall assessment scheme. For example, a participant used four assessment instruments with the following weightings and provided the following types of feedback: Final exam (60%) providing type 0, tests (20%) providing type 1, quizzes (10%) providing type 1 and homework (10%) providing type 2. For this participant, their three feedback measures were: 60 for type 0, 30 for type 1 and 10 for type 2.

Along with this work, an analysis was conducted on comments left by participants concerning both the assessment instruments they used, covered in the previous question, and the visible feedback they provided (see Appendix F). These comments were left in response to question 18 where participants were invited to leave any comments concerning their answers to questions 16
and/or 17. To prepare for this analysis, the comments were grouped into three categories: those associated with question 16 about participants’ course assessment instruments, those associated with question 17 about associated feedback practices and a third group for any miscellaneous comments. Following this, each group of comments was analyzed by summarizing emergent themes (i.e. thematic analysis; e.g. Braun and Clarke, 2006). Finally, any overarching themes were summarized and reported along with the results.

8.6.4 Use of Invigilation

Study I - R3: How are instructors using invigilation in their FO mathematics courses?

Use of invigilation was considered as a factor in determining any differences in how assessment instruments were weighted and how feedback was used. First, stacked column charts were created to compare assessment schemes of those using to those not using invigilation. Second, descriptive statistics provided summaries of the percentage of participants using invigilation for at least one instrument, the percentage of each instrument that was and was not invigilated as well as the associated average weighting.

Finally, analysis was conducted to investigate whether there were any differences in how feedback was used in the invigilated and non-invigilated groups. Results from this analysis were used to determine whether claims of an emphasis on formative assessment practices in non-invigilated FO courses were upheld (Trenholm, 2007a), with the expectation of a greater emphasis on feedback indicating an emphasis on formative assessment practices (e.g. Yorke, 2003).

8.6.5 Availability and Use of Professional Development

Study I - R4: What kind of professional development opportunities are FO mathematics instructors receiving for their courses?

Descriptives statistics were provided regarding six different types of PD opportunities. They showed percentages regarding how many participants used, had available but did not use, and did not have available. In addition, the numbers of days of PD participants engaged in was provided. Prior to compiling descriptive statistics, days of PD were converted from the raw string values to numerical values. In contrast with feedback timing, raw values for PD did not follow a direct scale conversion. Instead, ‘workday’ time was considered by equating one week
to 5 days and 1 day to 6 hours. For example ‘2 wks’ was equated to 10 days and ‘45 hrs’ was equated to 7.5 days. However, time intervals were treated as with feedback timing.

8.6.6 Approaches to Assessment Practice as Measured Using the S&B Framework

**Study I - R5:** Using the findings of Samuelowicz and Bain’s (2002) ‘Identifying academics’ orientations to assessment practice’ study as a framework, how are FO mathematics instructors approaching assessment in their courses?

This section details the operationalization of the S&B framework. This is followed by a summary of how the S&B data is subsequently analyzed using the framework.

**Operationalizing the S&B Study Framework**

A novel approach for analyzing instructors’ approaches to assessment was created by operationalizing the S&B study framework. This was accomplished by considering the six dimensions as six questions with possible responses corresponding to the different beliefs associated with each dimension. In the survey, this meant that participants were asked to identify, for each dimension, the belief that best represented their view. To do this, two changes were made to the original framework. First, the question terminology was modified to replace the original wording of ‘believes’ or ‘views’ to ‘in your approach’. As discussed in the literature review, this was not considered a major deviation from the original aim of the S&B study (see 5.2.6 for further discussion). Second, to prevent participants from detecting any patterns and alleviate any response order effects, a random number generator was used to scramble the order of the selections for each question (see Appendix G; Krosnick & Presser, 2010). From this, overall survey responses were then re-ordered into their original order and finally mapped back onto the S&B matrix. This provided an overall dimension by dimension picture of orientations, on the KR to KC continuum.

**Construction of the S&B Summated Scale**

Using the S&B framework presented in the literature review (see Table 1) a novel summated scale was constructed which presented a single quantitative measure of each individual participant’s approach to assessment practice. The idea arose from observing the structure of the framework where it appeared to this researcher that the way the approaches to assessment were being ranked for each dimension may possibly be translated into a quantitative measure. Such a scale measure was also considered desirable to be able to conduct quantitative analysis in
relation to assessment weightings. In this sense, the S&B summated scale was proposed and, as will be discussed in the next section, its validity will be investigated using the ATI.

This measure was calculated using the S&B number ranking corresponding to different orientations (see Table 9). For each question/belief dimension, where a single response mapped onto a single orientation, the number rank was used as a score. However, because all of the S&B dimensions contained at least one response that crossed two to three orientations the average of the number ranks was used as the score. For example, in the first belief dimension (‘Nature & structure of knowledge’) consisting of three possible responses, the score for the first response (see Table 9) was one as it mapped directly onto the first orientation with rank of one (‘1: Reproducing Bits of knowledge’). The second possible response mapped onto the next two orientations (‘2: Reproducing structured knowledge’ and ‘3: Applying structured Knowledge’) and the score was the average of the two number ranks (i.e. average of 2 and 3) or 2.5. The final possible response mapped onto the next the final three orientations (‘4: Organising subject knowledge’, ‘5: Transforming discipline knowledge’ and ‘6: Transforming conceptions of the discipline/world’) and the score was the average of these three number ranks (i.e. average of 4, 5 and 6) or 5. Depending on the responses selected by each participant for each question, these associated scores were summed so as to provide a measure of an individual participant’s approach to assessment practice. This measure ranged from a possible score of 7.5 to 31.5 corresponding to approaches to assessment from KR to KC.
### Table 9: S&B matrix showing scores used to operationalize the findings

<table>
<thead>
<tr>
<th>Orientations to assessment practice (with associated ranks)</th>
<th>Belief Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Reproducing Bits of knowledge</td>
<td>1</td>
</tr>
<tr>
<td>2: Reproducing structured knowledge</td>
<td>2.5</td>
</tr>
<tr>
<td>3: Applying structured knowledge</td>
<td></td>
</tr>
<tr>
<td>4: Organising subject knowledge</td>
<td></td>
</tr>
<tr>
<td>5: Transforming discipline knowledge</td>
<td></td>
</tr>
<tr>
<td>6: Transforming conceptions of the discipline/world</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis: S&B Framework

Two principal stages of analysis were performed. The first stage used the total number of participants selecting each response for each question. Following this, each of the six S&B questions, representing each of the six S&B belief dimensions, were analyzed one at a time to determine which end of the continuum, KR or KC, the majority of participants were found (as delineated by the red line shown in Table 9) - that is, whether the majority of the participants selected responses in the first three orientations (the KR half) or the last three orientations (the KC half).

Along with this first stage, participants’ comments provided with the S&B question responses (see Appendix K) were analyzed. To prepare for this analysis, the comments provided with each S&B question were grouped according to which of the three to four responses – associated with
the KR extreme, mid-range orientation(s) and KC extreme of the continuum – they accompanied. That is, for each of the first five S&B questions which each had three possible responses, comments were placed into one of three groups according to which S&B question response it accompanied. For the last question, which had four responses, comments were placed into one of four groups according to which question response it accompanied. For example, for the first S&B question, five comments were left by five participants who selected the response associated with an orientation to assessment at the KR extreme of the continuum, one comment was left by one participant who selected the response associated with the mid-range of the continuum and one comment left by one participant who selected the response associated with the KC extreme of the continuum. Following this, for each S&B question, these groups of comments were analyzed by summarizing any emergent themes (i.e. thematic analysis; e.g. Braun and Clarke, 2006). And finally, for each S&B question, any overarching themes were summarized and then, for all S&B questions, any overall emergent themes were summarized. These results are reported along with the results in the first stage of analysis.

In the second stage, as previously detailed, individual participant S&B measures were calculated and then the distribution was investigated for normality – both as a basis for arguing the validity of the scale (Kemp & Grace, 2010) and to determine whether parametric tests were to be used for the coming analysis involving the S&B approach measure. The third and final stage investigated, using correlational analysis, whether there was any relationship between the new S&B measure and any of the feedback measures.

8.6.7 Approaches to Teaching as Measured by the ATI

Study I - R6: How are FO mathematics instructors approaching teaching in their courses as measured by Prosser and Trigwell’s (2004) Approaches to Teaching Inventory (ATI)?

As per Trigwell and Prosser (2004), questions were presented in the order specified (i.e. ATI01, ATI02, ATI03…). The only modification made was to the wording for question ATI06 where, in order to fit the nature of FO instruction, the wording was changed from ‘I set aside some teaching time’ to ‘I plan my teaching’ (see Appendix A for original questions). From participants’ answers, two ATI measures (Prosser & Trigwell, 2006) were derived by adding the values, ranging from one to five, corresponding to the eight specific likert-style questions for each of the two scales (CCSF and ITTF; see Appendix A). These measures presented a gauge of how instructors approached their teaching. The first scale (i.e. CCSF) measured to what degree
an instructor’s approach was student-focused and ‘aimed at conceptual change’ (Stes et al., 2010, p.60). The second scale (i.e. ITTF) measured to what degree an instructor’s approach was teacher-focused and ‘geared towards information transmission’. Both scales provided a measure from a minimum of 8 to a maximum of 40 of how instructors were oriented to the respective scale description.

A normality test was also conducted on both ATI measures along with descriptive statistics (i.e. mean and standard deviation) calculated for each of the scale measures. Following this, as with the S&B measure, correlational analysis was used to investigate whether there was any relationship between the ATI measure and any of the feedback measures. Finally, comments left by three participants concerning the ATI questionnaire were summarized.

**8.6.8 An Investigation of the Relationship Between the S&B and ATI Measures**

*Study I - R7: How do findings in question six relate to findings in question five?*

The modified S&B statements generally asked ‘In your approach to assessment, which of the following descriptions best describes how you...’ with possible responses generally corresponding to the participants perceptions of actual use. Similarly, the ATI considered instructor intention and strategy together. That is, the two frameworks were assumed to measure a similar underlying construct. Therefore the concurrent validity of the new S&B measure was investigated using the ATI, as a valid established psychometric instrument. This was accomplished by testing the correlation of the S&B measure with each of ATI scale measures. Conceptually, a degree of validity was seen to be established if the CCSF measure was significantly and positively correlated with the S&B measure and the ITTF measure was significantly and negatively correlated with the S&B measure.

**8.6.9 Use of the Study Frameworks to Investigate the Use of Invigilation, a Variety of Instruments, Quizzes and Discussion**

*Study I - R8: Are there any statistically significant differences in any of the study measures based on usage of invigilation, a greater variety of instruments, quizzes or discussion (the latter two as weighted instruments)? When used, is the weighting of either quizzes or discussion related to any of the study measures? Do these findings support prior claims and findings?*
Guided by claims and findings in the literature, two types of analyses were conducted to investigate the relationship of invigilation, number of assessment instruments, quizzes and discussion (with the latter two as weighted assessment instruments) to each of the six study measures. In the first analysis, a sample population comparison was used to investigate if any of the study measures differed significantly based on whether invigilation, a higher or low number (e.g. 1-3 and 4+) of assessment instruments, quizzes or discussion were used. In the second, a correlational analysis was used to investigate the relationship between quiz and discussion emphases and the study measures. These tests were intended to explore associations between variables.

8.7 Validity and Reliability

As a first step in answering the research questions, FO mathematics instructors were sampled using a survey instrument that was distributed online. This data was then analyzed using the three study frameworks discussed earlier. This section begins by first discussing the validity of three study frameworks. Though the validity and reliability of the ATI was discussed earlier in the literature review, here these issues are discussed in the context of how the ATI was used. Following this, the psychometric validity and reliability of the survey instrument along with the data collection process is covered.

Issues of Validity: Feedback Framework

While the literature (see 6.2.3 in the literature review) formed the basis for how this framework was created, there were at least three validity issues. First, the framework relied on generic de-disciplined and limited mathematics-specific findings. That is, apart from generic findings, it was unclear how previously cited feedback findings for statistics and geometry applied, for example, to calculus. Second, the ‘computer’ or ‘lecturer-generated hints or comments’ description were open to interpretation. In particular, as discussed in the literature review on assessment, if these ‘hints or comments’ were directed at the ‘self level’, this feedback was considered to be ineffective in enhancing learning (e.g. Hattie & Timperley, 2007). For example, if participants were referring to ‘praise’ or ‘encouragement’ when referring to any ‘hints or comments’ this would not be considered by the literature as ‘rich’ feedback that enhances learning. In short, in terms of potential effects on learning, type 2 feedback may have been operating like type 0 feedback. Further research needs to specify this feedback as, for example, directed at students’ mathematical thinking. Third, the feedback measures were
premised on the assumption that assessment weighting (and not simply ‘assessment’) directed student learning. This assumption, as discussed in the literature review, had received little attention in the current research and therefore needs testing.

**Issues of Validity: S&B Framework**

While the validity of the original S&B study findings may be argued based on the strong relationship they had to prior research (Samuelowicz & Bain, 2002), the primary issues concerning this study related to how the framework was being operationalized. As such three areas will be discussed: how the questions were created, how the S&B matrix was used for analysis and how the S&B measure was created.

First, while questions were created, almost word for word, using the descriptions of the orientations in the original study, there was one significant change made in the process of operationalizing the original framework that may be considered a threat to validity. This change involved substituting the original wording of ‘believes’ or ‘views’ with ‘in your approach’. Though, as previously discussed, this was not considered a major deviation from the original aim of the S&B study (see 5.2.6 for further discussion), these changes, nevertheless, represented a change from the original wording that may be considered a threat to validity.

Second, given the analysis considered how the number of responses split either towards KR or KC (as delineated by the red line shown in Table 9), in four of the questions (i.e. questions 2 and 4-6) the response straddled both sides of the line. As a result, if a large number of participants selected any of these four responses, it was unclear which way participants split for that dimension. This was an obvious limitation effecting the validity of this approach to analysis.

Third, perhaps the greatest threat to validity concerned the ordinal nature of the S&B scores and the issue of integrating this data into a summated scale. The primary concern regarded the issue of ‘trade-offs’ where ordinal variables that were not ‘strongly positively’ correlated lead to the cancellation of effects when the scores were summed across variables (Kemp & Grace, 2010). For example, after a participant submitted an answer for the first question, while answering another question, he/she may have felt the need, to be able to answer both questions accurately, to change the answer to the first question (i.e. ‘trade-off’). In consideration of this issue, there were at least three characteristics that were said to alleviate any such concerns (Kemp & Grace, 2010). First, if the summated scale measures were normally distributed this suggested a possible underlying interval scale that justified a summated scale measuring the psychological processes.
addressed in the S&B questions. Second, if the question presentation was such that participants were given the opportunity to consider all questions and responses together (vs. singly), this may potentially alleviate any ‘trade off’ issues. Third, summation was considered appropriate if the model adequately described the ‘processes and phenomena’ (p. 407) being researched. The first characteristic will be considered in the results section, based on the nature of the survey and the S&B framework. The second characteristic was assumed to be adequately addressed given the survey presented all questions and responses together with clear directions for participants to ‘take the best’ response and, if necessary, provide comments (p.408). The third issue, that the nature of the original S&B study and findings presented a framework model that adequately describes mathematics assessment practice was a little more difficult to address. The vast majority of the S&B study participants taught pure applied disciplines (e.g. three from physiotherapy, five from architecture...) with four out of 20 from a single pure hard discipline (i.e. chemistry). While the disciplinary groupings studied were somewhat aligned, they were not mathematics and therefore it was possible that the study findings and the frameworks they presented were limited in their application to the thesis context. While the validity of the S&B approach measure will be checked using the ATI, this new measure requires further testing in the mathematics instructional context – particularly to investigate whether it needs to be adapted for this disciplinary context.

One final issue concerned the validity of how this measure might be used in understanding instructors’ approaches to assessment practice. That is, does the summated scale measure actually mean anything? The seventh research question explored this issue by using the ATI (Trigwell & Prosser, 2004) as a valid established psychometric instrument to compare with the S&B results. This will be discussed later in the results section.

**Issues of Validity: Use of ATI**

One major caveat concerned the use of the ATI for comparative purposes. The ATI was mainly intended ‘for the analysis of relations between approaches to teaching and other elements of the teaching-learning environment perceived by the *same teacher* in the *same context*’ (italics mine; Trigwell & Prosser, 2004, p.421). Limiting this study’s findings, the current use of the ATI departed from these expectations when similar teachers (mathematics instructors) in similar contexts (tertiary FO mathematics courses) were measured and compared. That is, for example, the same instructor may not have had the same approach to teaching a statistics course as he did
teaching a calculus course. A similar limitation regarding use of the S&B measure might be argued.

**Issues of Validity and Reliability: Survey Instrument**

According to Fink (2003), there were at least four validity and five reliability criteria concerning any survey instrumentation. Concerning validity, the first two criteria – face and content validity – were addressed in the pilot phase through iterative reviews by both F2F and FO instructors. For example, the online survey was initially administered to a few instructors from the target demographic to investigate if there were any related issues. Following this, only concurrent criterion validity was explored (construct validity of the survey was not investigated) by comparing S&B measures to ATI measures, as an established psychometric instrument.

Next, only one of the five reliability criteria was explored (test-retest, alternate-form and inter and intraobserver reliability were not investigated). This criterion, internal consistency, was investigated using Cronbach’s alpha for both the S&B measure and the ATI. First, a reliability coefficient of 0.450 was calculated for the S&B questions, which was considered to be low. However, three characteristics of the S&B scale were expected to contribute to poor reliability: when a small number of questions were used or the scale was multidimensional (Cortina, 1993) and when the number of responses available for each question was only between two and four (Preston & Colman, 2000). In balance, while the low reliability coefficient suggested the internal consistency of the S&B measure was questionable, the characteristics of the scale suggested the measure may be reliable. Second, the alpha coefficients for the two ATI scales were 0.789 (CCSF) and 0.505 (ITTF). This suggested the inventory had good statistical validity for the CCSF scale but questionable validity for the ITTF scale (Trigwell & Prosser, 2004). The one caveat was that the lower alpha value for the ITTF scale, as compared to the CCSF scale, was somewhat in line with previous general (0.66 vs. 0.74 for n=1023, Prosser & Trigwell, 2006) and mathematics-specific findings (0.707 vs. 0.8, n=177 and 176 respectively, Andersen, 2011). However, in summary, this cast some doubt on the significance of any findings related to the ITTF scale.

Finally, a related form of reliability, discussed earlier in the methodology, concerned some of the terminology used in the survey. While pilot work sought to address some issues of terminology (e.g. UK ‘module’ vs. US ‘course’), survey comments revealed that use of the term ‘invigilation’ was unfamiliar to one US participant who was more familiar with the term ‘proctoring’. Another issue concerned the lack of definitions for different assessment
instruments – in particular, those specified in the survey (i.e. homework, quizzes, tests, midterm exam, final exam, individual project, group project, portfolio, group work, journal and discussion). Terms such as ‘final exam’ (as a well-known assessment component in mathematics) and ‘discussion’ (as well-known assessment component in FO instructional contexts) may have been commonly understood. However, for other instruments, the distinctions were not clear. For example, in the context of this study, it was assumed quizzes were similar to tests but shorter and administered more frequently while homework was similar to quizzes but longer and intended for completion ‘at home’. Therefore, an assumption was made that participants shared a common understanding about the terminology and nature of each assessment instrument. However, given this may not be the case, it presented a limitation to the validity of the survey. This limitation was perceived to be greatest concerning ‘quizzes’, where the actual number of users might have been much greater if some participants used the term ‘tests’ to refer to the same kind of assessment instrument (or vice-versa). Further limitations were avoided when much of the analysis relied on the kind of feedback associated with each instrument, rather than simply its label, as the basis for distinguishing assessment instruments.

**Issues of Validity and Reliability: Use of an Online Survey for Data Collection**

While online surveys were viewed, for example, as a convenient and efficient means of collecting data (Evans & Mathur, 2005), several concerns relating, for example, to sampling validity and non-response bias (Duda & Nobile, 2010) have been raised. There were at least three issues concerning the validity and reliability of the online survey used for this study. First, the online survey used for this study involved nonprobability convenience sampling where participants were selected based on prior contact with this researcher. Second, the online survey also involved snowball sampling where some research contacts were asked to pass the survey link onto other potential participants. Third, the Bristol Online Survey (BOS) system, used to administer this survey, allowed participants to skip questions asking ‘check all that apply’. In these instances, the system output produced ‘not applied’ when this may not be true. The first two issues limited the external validity of the results while the last limited the reliability of the survey instrument.

These limitations were not considered to invalidate the first study findings. Many limitations, particularly regarding the S&B and feedback frameworks, presented some caution concerning generalizability. However, perhaps more so, they presented direction for future research.
8.8 Summary

- An initial qualitative pilot study lead to selecting an online survey methodology.
- The ATI, the S&B study framework and the feedback framework were selected for the analysis. A questionnaire was constructed for a pilot survey. Based on the pilot survey feedback, a survey questionnaire was finalized.
- Potential participants were contacted resulting in 70 completed surveys.
- Following the operationalization of the frameworks, a total of six study measures were used for the analysis: the ATI consisting of the CCSF and ITTF measures, the S&B summated scale measure as well as three feedback measures. Using the S&B framework and the six study measures the survey data was displayed and analysed.
- Related issues of validity and reliability were discussed and knowledge claims were qualified.
9. Study I Results and Discussion

The results for the first study are divided into three parts (Parts I, II and III) and detail findings as they pertain to each of the first eight research questions (see Table 10). In Part I, participant demographics are first provided. Next, Part II outlines results as they pertain to the first four questions (i.e. R1, R2, R3 and R4) which explore the actual teaching and assessment practices of FO mathematics instructors. The first three of six study measures are also introduced, relating the kind(s) of feedback provided by participants to their assessment weighting scheme. Finally, Part III outlines results as they pertain to the remaining four questions (i.e. R5, R6, R7 and R8) which explore the teaching and assessment approaches of FO mathematics instructors, particularly as they relate to the three approach measures. With the majority of participants from US institutions of higher education, findings relate to both assessment practice in general and the FO context in particular.

Table 10: Study I research questions and corresponding analysis

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study I: Part II</td>
<td></td>
</tr>
<tr>
<td>R1. What instruments are FO mathematics instructors currently using to assess their students? How are these weighted?</td>
<td>Descriptive by instrument and participant</td>
</tr>
<tr>
<td>R2. How are instructors using feedback in their FO mathematics courses?</td>
<td></td>
</tr>
<tr>
<td>R3. How are instructors using invigilation in their FO mathematics courses?</td>
<td>Descriptive by instrument and participant</td>
</tr>
<tr>
<td>R4. What kind of professional development opportunities are FO mathematics instructors receiving for their courses?</td>
<td>Descriptive by type and days of PD</td>
</tr>
<tr>
<td>Study I: Part III</td>
<td></td>
</tr>
<tr>
<td>R5. Using the findings of Samuelowicz and Bain’s (2002) ‘Identifying academics’ orientations to assessment practice’ study as a framework, how are FO mathematics instructors approaching assessment in their courses?</td>
<td>Descriptive by question Analysis using S&amp;B framework Correlation with feedback measures</td>
</tr>
<tr>
<td>R6. How are FO mathematics instructors approaching teaching in their courses as measured by Prosser and Trigwell’s (2004) Approaches to Teaching Inventory (ATI)?</td>
<td>Descriptive by subscale and question Correlation with feedback measures</td>
</tr>
<tr>
<td>R7. How do findings in question six relate to findings in question five?</td>
<td>Correlation</td>
</tr>
<tr>
<td>R8. Are there any statistically significant differences in any of the study measures based on usage of invigilation, a greater variety of instruments, quizzes or discussion (the latter two as weighted instruments)? When used, is the weighting of either quizzes or discussion related to any of the study measures? Do these findings support prior claims and findings?</td>
<td>Sample population differences with usage and correlation with weighting by study measures</td>
</tr>
</tbody>
</table>
Presentation of Results and Use of Analysis

In general, the results are presented from two perspectives: assessment instrument and individual participant. First, an assessment instrument perspective is provided by means of tabular displays (e.g. see Table 12). Tables provide descriptive statistics for each of the assessment instrument types (ordered from the top to bottom from those most to least used). Second, an instructor or participant perspective is provided by means of stacked column graphs (e.g. see Figure 2). These graphs colour-code the usage and weighting of assessment instrumentation for each individual participant. Two types of stacked column charts are used. The first displays participants’ overall assessment instrument usage and weightings by assessment instrument categories used in the survey questionnaire (i.e. homework, quizzes, tests, final exams...; e.g. see Figure 2). To improve readability in these graphs, homework will always be coded in a contrasting dark blue colour. The second displays participants’ overall assessment instrument usage and weightings by considering, for each participant, the sum total of assessment instrument weighting associated with each of three feedback types (i.e. sum of the weightings of instruments using type 0 feedback, using type 1 feedback and then using type 2 feedback or what is essentially a participants’ three feedback measures; e.g. see Figure 3). In each stacked column chart participants are sorted as detailed on the horizontal axis labels. All graphs (including scatter plots in Part III) are constructed in Excel. They provide a visual description of the data which is used to search for any discernible patterns. Finally, tables are summarized with particular attention to high and low measures.

Descriptive statistics and analysis are conducted using PASW (formerly SPSS statistics). The focus of analysis is on sample population differences and correlation. Differences in sample populations here concerns testing to see if there are any significant differences in the ‘use’ and ‘non-use’ of a particular assessment instrument (e.g. those that do and those that do not use quizzes). Correlational analyses here concerns testing to see if the weighting of a particular assessment instrument is significantly correlated with one or more of the study measures (e.g. Is the weight given to quizzes in the overall assessment scheme correlated with the S&B measure?). In either case, prior to any analysis, normality tests are conducted to determine whether parametric or non-parametric tests should be used.
9.1 Part I: Participant Demographics

A total of 70 participants completed the survey. Of those participants, four failed to provide all of their assessment specifics and were excluded from any analysis related to assessment specifics. Other issues with missing or problem data are detailed, with actions taken, in Appendix E. Table 11 provides a breakdown of overall demographic characteristics of participants.

In summary, the data indicates that the majority of participants are experienced mathematics instructors, most of whom working full-time in either 2 or 4 year US institutions where their work is mostly focused on teaching. Most have taught online for one to ten years and just over half have only taught one FO mathematics course. Finally, most participants select either undergraduate calculus or statistics as the course context for answering the survey questions.
Table 11: Participant demographics

<table>
<thead>
<tr>
<th>Demographic Category</th>
<th>Characteristic</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin (based on survey link only)</td>
<td>US</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Outside US</td>
<td>24</td>
</tr>
<tr>
<td>Role in academia</td>
<td>Mostly research (education focus)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mostly research (pure mathematics/statistics focus)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mostly teaching</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>About the same amount of research and teaching</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Other (specify): Administrator or professional tutor also teaching FO</td>
<td>6</td>
</tr>
<tr>
<td>Status</td>
<td>Full-time staff/faculty</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Part-time staff/faculty</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Neither (unknown)</td>
<td>1</td>
</tr>
<tr>
<td>Institution</td>
<td>Traditional ‘bricks and mortar’ bachelor degree granting (e.g. university, 4 yr college)</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Traditional ‘bricks and mortar’ non-bachelor granting (e.g. 2 yr North American community college)</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Open/Online university/college</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Other (specify): Institute of Technology Ireland</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Not Specified</td>
<td>3</td>
</tr>
<tr>
<td>Years of experience teaching $F2F$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1-5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>6-10</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>11-15</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>16+</td>
<td>63</td>
</tr>
<tr>
<td>Years of experience teaching $FO$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>&lt;1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1-5</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>6-10</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>11-15</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>16+</td>
<td>1</td>
</tr>
<tr>
<td>Number of different FO courses taught (in the last three years)</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Survey course focus</td>
<td>FO pre-calculus</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>FO calculus I, II or III</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>FO introductory statistics</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>FO online advanced statistics</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Other (see Appendix H)</td>
<td>31</td>
</tr>
<tr>
<td>Survey course focus level</td>
<td>Undergraduate</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Graduate</td>
<td>7</td>
</tr>
</tbody>
</table>
9.2 Part II: Teaching and Assessment Practices

This part of the first study results covers the first four research questions. The focus is on investigating the assessment practices in a single FO course for each participant. This is accomplished by using supporting tables and figures relevant to each question. In addition, for two questions, sample population differences are explored using appropriate statistical tests.

Overall, participants’ FO assessment practice appears to mirror what is known of F2F practice where summative-style instruments are emphasized. Findings also provide support for some claims and findings in the literature, such as a decline over several years in the use of discussion, as well as some possible insights into current FO mathematics instructional and professional development practices.

Study I - R1: What instruments are FO mathematics instructors currently using to assess their students? How are these weighted?

In the current research literature, there is little known about tertiary mathematics assessment practice and even less known about FO mathematics assessment practice. While overall findings suggest the traditional F2F emphasis on summative-style instruments continues (e.g. Iannone & Simpson, 2012a), evidence of other more ‘non-traditional’ instruments (e.g. discussion) is evident and wide variability in weighting is apparent in many of the instruments.

Instrument Perspective

Table 12 provides information regarding the number of participants using each of 12 assessment instruments (including one ‘other’ category) as well as statistics on instrument weighting. Generally, all instruments that participants report using count towards the overall course grade. However, in at least two cases, participants report using instruments that do not count in their overall assessment scheme. It is not clear if these are, for example, gateway activities that need to be completed before a student can proceed to instruments that do count.
Table 12: Assessment instruments used with descriptive statistics

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Total Number Using (of n=66)</th>
<th>Weighting (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Homework</td>
<td>55</td>
<td>28</td>
</tr>
<tr>
<td>Final Exam</td>
<td>48</td>
<td>33</td>
</tr>
<tr>
<td>Tests</td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td>Quizzes</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Mid-Term Exam</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Individual Project</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Other Assessment*</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Group Project</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Group Work</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Journaling</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Portfolio</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

*Includes one participants’ instance of extra quiz work and two participants’ instances of extra homework, all with different assessment and/or feedback specifics than the initial recorded quiz/homework instance.

In terms of weighting, most FO mathematics instructors tend to emphasize summative-style instruments (i.e. final exam, tests and mid-term exam), mirroring the emphasis in F2F courses (e.g. Burn et al., 1998). Tests and final exams, for example, have an average weighting of 42% and 33%, respectively. Similarly, the number of participants using homework (55 of 66; 83%) suggests the value these participants place on this kind of instrument. However, most instrument weightings show high variability. Homework, notably, has very high variability (SD = 26.4). In addition, there is also some evidence of ‘non-traditional’ instruments in use, most notably the use of discussion and individual projects. Finally, two types of group-oriented instruments (group projects and group work activities) as well as two types of individually-oriented instruments (journaling and portfolio work) do not appear to be used very often. Given the low participant numbers found in the last four instrument categories, the remaining study focus will be on the top eight identified instruments (see Table 12).

Finally, the use of discussion and quizzes as weighted assessment instruments (and two foci of the thesis) are summarized and related to the literature. Fewer than half (26 of 66 participants or 39%) of the participants are found to use discussion as a weighted (mean weighting of 10%) assessment component in their FO course. In relation to prior findings, the average weighting is about the same but the percentage of instructors using discussion is considerably lower (78 percent using weighted at 7 percent, Galante, 2002; 70 percent using weighted at 13 percent, Trenholm, 2007a). Moreover, as discussed in the literature review, it was found, using
comparable data (i.e. from SUNY instructors) over a five year period, that hard pure courses (i.e. mathematics) are moving away from a community orientation. Consistent with this, recent qualitative research at the school-level in Australia finds teaching of mathematics at a distance is a primarily ‘one-on-one’ experience with little use of collaboration (Lowrie & Jorgensen, 2012). This study presents evidence consistent with these findings where a decline in usage as well as weighting is found when comparing SUNY participant data for this thesis with data from the 2007 study (see Appendix I for a breakdown of courses, Trenholm, 2007a). As shown (see Table 13), on average, compared to 2006 courses, 18.6% fewer 2010 course instructors use discussion as a weighted component of their assessment scheme and, when used, weight it 3.1% less. A test is run comparing, in 2006 and 2010, overall participants’ use of discussion as a weighted assessment instrument with those not discussion considered as 0% weighting. On average, the 2010 discussion weighting (Mdn=2.00) is significantly less than the 2006 discussion weighting (Mdn=10.00), U=786, Z= -2.5, p < 0.05, r= -.25.

Table 13: Use and weighting of discussion for SUNY participants

<table>
<thead>
<tr>
<th>Course Type</th>
<th>2006*</th>
<th>2010</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants Using</td>
<td>Average Weighting</td>
<td>Participants Using</td>
</tr>
<tr>
<td>Calculus</td>
<td>75% (6/8)</td>
<td>13.3%</td>
<td>50% (6/12)</td>
</tr>
<tr>
<td>Statistics</td>
<td>61.5% (8/13)</td>
<td>11.3%</td>
<td>45.5% (5/11)</td>
</tr>
<tr>
<td>Other</td>
<td>71.8% (28/39)</td>
<td>13.0%</td>
<td>57.1% (8/14)</td>
</tr>
<tr>
<td>All</td>
<td>70% (42/60**)</td>
<td>12.9%</td>
<td>51.4% (19/37)</td>
</tr>
</tbody>
</table>

*Data collected in 2006. **The original study considered all sections taught by the same instructor, including multiple sections. The data used here considers multiple sections taught by the same instructor only once.

Lastly, a little more than half (35 of 66 or 53%) the participants use quizzes as a weighted (mean weighting of 20%) component of assessment in their FO course. This appears to contrast with expectations, discussed in the literature review, that the use of quizzes is a central characteristic of FO instruction (Engelbrecht & Harding, 2005a; Greenberg & Williams, 2008). In relation to prior 2006 SUNY data (see Table 14), the overall average quiz weighting in 2010 is about the same (+1.5%) but the percentage of instructors using quizzes is somewhat higher (+8.1%) – with the use of quizzes in calculus, in particular, increasing sizeably in both usage and weighting. However, no statistically significant difference is found when comparing, in 2006 and 2010, overall participants’ use of quizzes as a weighted assessment instrument with those not using
quizzes considered as 0% weighting. While differences appear high in Calculus, sample sizes are too small to test for any significant difference.

Table 14: Use and weighting of quizzes for SUNY participants

<table>
<thead>
<tr>
<th>Course Type</th>
<th>2006*</th>
<th>2010</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants Using</td>
<td>Average Weighting</td>
<td>Participants Using</td>
</tr>
<tr>
<td>Calculus</td>
<td>50%</td>
<td>12.8%</td>
<td>66.7%</td>
</tr>
<tr>
<td></td>
<td>(4/8)</td>
<td></td>
<td>(8/12)</td>
</tr>
<tr>
<td>Statistics</td>
<td>30.8%</td>
<td>31.5%</td>
<td>36.4%</td>
</tr>
<tr>
<td></td>
<td>(4/13)</td>
<td></td>
<td>(4/11)</td>
</tr>
<tr>
<td>Other</td>
<td>46.2%</td>
<td>21.6%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>(18/39)</td>
<td></td>
<td>(7/14)</td>
</tr>
<tr>
<td>All</td>
<td>43.3%</td>
<td>21.8%</td>
<td>51.4%</td>
</tr>
<tr>
<td></td>
<td>(26/60**)</td>
<td></td>
<td>(19/37)</td>
</tr>
</tbody>
</table>

*Data collected in 2006. **The original study considered all sections taught by the same instructor, including multiple sections. The data used here considers multiple sections taught by the same instructor only once.

**Individual Participant Assessment Practice Perspective**

Figure 2 presents a method of displaying and comparing participants’ assessment schemes. It provides a visual perspective of the assessment practices for each of the 66 participants (ordered left to right by homework and then final exam weighting from highest to lowest). The graph, for example, shows the prevalent use of homework and final exams with associated weightings. For homework, in particular, weightings can be seen to vary considerably across participants. Note the one column above the 100% mark is assumed to represent possible bonus points. A practice that is not entirely uncommon in US higher education. Also, as shown, another column is below 100%. Here it is assumed to be a participants’ mistake.
Figure 2: Individual participants’ assessment practices (Note: For this and the following stacked column charts, where the stacked column is above 100% it is assumed to represent possible bonus points and where the column is below 100% it is assumed to be a participants’ mistake)
Study I - R2: How are instructors using feedback in their FO mathematics courses?

There is an emphasis in the current literature on feedback as a critical distinguishing characteristic of assessment instruments, particularly with regards to its value in student learning (e.g. Taras, 2005). So, as outlined in the methodology, a definitional framework is created that divides feedback, in order of quality, into three types. Any assessment instrument providing feedback consisting only of a grade is considered type 0 and given a feedback score of zero. Those providing any or no type 0 feedback and the answer or full solution as feedback are considered type 1 and given a score of one. Those providing any or no type 0 and/or type 1 feedback with hints or comments are considered type 2 and given a score of two. For example, providing the full solution and hints are considered type 2.

Using this framework, the assessment instrumentation used by each participant and each participant’s overall assessment scheme is classified by the associated feedback used. Individual assessment instruments are classified according to the associated feedback type and this information is used to calculate an average feedback measure for each type of assessment instrument (see Table 15). In addition, individual participants’ assessment practices are summarized by calculating the sum of the instrument weightings for each type of feedback used. This then is displayed using a stacked column chart (see Figure 3).

The analysis provides a new way of viewing each instrument. In particular, findings appear to provide an indication of overall participant intentions regarding the degree to which each instrument is intended for formative or summative purposes. Additionally, the graph provides a new way of visualizing overall participant assessment schemes.

Instrument Perspective

To begin, feedback at the instrument level is investigated. First, the average feedback score for each instrument is calculated. Again, as per the literature, based on the feedback framework, this average score is used to consider the potential effect of an instrument on learning. Second, for each instrument, the number of participants using each of the three kinds of feedback is detailed (i.e. with the sum equal to the total number of participants using the particular instrument). For example, with the use of tests, 8 participants provided type 0 feedback, 17 provided type 1 and 18 provided type 2, for a total of 43 participants. Third, timing of feedback (in days) is also investigated. To begin, raw data, for example, in hours or weeks, is converted into days with the average used for time ranges (e.g. ‘3-5 days’ is treated as 4 days) and ‘less
than \( x \)’ answers considered as the average between 0 and \( x \) (e.g. ‘less than 1 week’ is treated as 3.5 days). The mean and standard deviation per instrument type is then calculated.

Table 15 and Table 16 provide descriptive statistics regarding the nature of feedback used by participants in their assessment practice. Referring to Table 15, the type of feedback associated with each assessment instrument is summarized in two ways. First, the average feedback score associated with each instrument is given in column three. Second, for each instrument, the number of participants using each feedback type is detailed in the remaining three columns.

Table 15: Quality of feedback as measured by the study framework

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Total Number Using (of ( n=66 ))</th>
<th>Average Feedback Score (2 d.p.)</th>
<th>Number of Participants Using each Type of Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Homework</td>
<td>55</td>
<td>1.73</td>
<td>1</td>
</tr>
<tr>
<td>Final Exam</td>
<td>48</td>
<td>0.52</td>
<td>30</td>
</tr>
<tr>
<td>Tests</td>
<td>43</td>
<td>1.23</td>
<td>8</td>
</tr>
<tr>
<td>Quizzes</td>
<td>35</td>
<td>1.26</td>
<td>1</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
<td>1.00</td>
<td>12</td>
</tr>
<tr>
<td>Mid-Term Exam</td>
<td>16</td>
<td>0.94</td>
<td>5</td>
</tr>
<tr>
<td>Individual Project</td>
<td>13</td>
<td>1.85</td>
<td>0</td>
</tr>
<tr>
<td>Other Assessment</td>
<td>7</td>
<td>1.50</td>
<td>1</td>
</tr>
</tbody>
</table>

The average feedback measure calculation indicates that the richest feedback appears to be associated with individual projects (1.85) followed by homework (1.73) with the poorest feedback appearing to be associated with final exams (0.52). With the possible exception of individual projects, these results are perhaps to be expected, with work pursued during the course associated with richer feedback whereas work administered at the end of the course is not. They also appear to support prior assumptions (Trenholm, 2007a) that ‘homework’ is primarily intended for formative purposes while ‘final exams’ are primarily intended to be summative.

Contrasting with the coarse-grained average feedback score associated with each assessment instrument, the final three columns of the table provide a finer-grained breakdown of feedback types used with each instrument. From this vantage point, a few observations are made. First, three instruments show definite tendencies displayed by the increasing numbers of participants from left to right or right to left. Both homework and individual project use tend in the direction of richer feedback while final exam use shows a tendency in the opposite direction (i.e. towards
poorer feedback). Second, with the use of tests, an almost equal number of participants provide type 1 and type 2 feedback. Third, the use of quizzes indicates a prevalent use of the intermediate type of feedback while discussion is polarized on both ends with little use of the intermediate feedback type. Fourth and finally, mid-term exams display no particular tendency with all feedback types almost equally used. Overall, based on the associated feedback provided, results indicate that different participants appear to intend to use the same named assessment instrument for different purposes (i.e. formative to summative). However, many instruments show clear tendencies to one primary purpose.

Following on from Table 15, Table 16 considers the number of times each instrument was assigned or administered as well as the feedback timing. The timing of feedback for the first seven instruments range from the shortest timing for quizzes (0.8 days) to the longest timing for individual projects (5.8 days), with wide variability noted for a number of instruments. Here again the use of homework is associated with wide variability.

Table 16: Timing of feedback

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Total Number Using (of n=66)</th>
<th>Frequency (Mean)</th>
<th>Timing of Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Homework</td>
<td>55</td>
<td>14.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Final Exam</td>
<td>48</td>
<td>1.1</td>
<td>4.8</td>
</tr>
<tr>
<td>Tests</td>
<td>43</td>
<td>6.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Quizzes</td>
<td>35</td>
<td>9.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
<td>8.9</td>
<td>4.6</td>
</tr>
<tr>
<td>Mid-Term Exam</td>
<td>16</td>
<td>1.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Individual Project</td>
<td>13</td>
<td>4</td>
<td>5.8</td>
</tr>
<tr>
<td>Other Assessment</td>
<td>7</td>
<td>5.7</td>
<td>4.3</td>
</tr>
</tbody>
</table>

**Individual Participant Assessment Practice Perspective**

Figure 3 displays individual participants’ assessment instrument weighting associated with each type of feedback (with participants ordered from left to right by weightings associated with a feedback score of 0 and then by 1). Overall, 26.6% of all assessment weightings are associated with use of type 0 feedback, 32.2% with type 1 and 41.3% with type 2. As shown most (48/66
or 73% of all participants) use a combination of two or more feedback types though a number report using type 2 feedback (12/66 or 18%) for all of their assessment instruments (vs. 3/66 or 5% using only type 1 or only type 2 feedback). In balance, participants appear to emphasize the use of type 2 feedback with most using a combination of two or more types of feedback.

**Analysis of Participants’ Comments**

Finally, in question 18, participants were invited to comment on their answers to questions 16 and/or 17, regarding the assessment instruments they use and the visible feedback they provide along with them (see Appendix F). The following paragraphs provide a summary of the fifteen participant comments that were left. They are separated into three categories largely according to which survey question is being addressed. First, comments directed at participants’ assessment schemes are summarized. Second, comments directed at participants’ feedback practices are summarized. Third, two remaining comments that are judged to be outside the previous categories are summarized. Comments generally clarify, and, in so doing, suggest different levels of complexity in participants’ assessment and feedback practices.

First, six comments clarify participants’ assessment schemes. They suggest these schemes, including the rationale behind them and the contribution they make to a student’s overall course grade, are complex. For example, one participant uses a small ‘participation’ grade to ‘leverage’ student engagement. Another requires students earn a passing grade on their invigilated midterm and final exams in order to pass a course (i.e. all other grades are otherwise inconsequential). Finally, one participant clarifies their assessment practice by detailing the number of attempts students are permitted in completing a particular assessment instrument and which of the associated grades ‘count’ towards their overall grade.

Second, seven comments relate to participants’ feedback practices. As with those referring to participants’ assessment schemes, these comments suggest that the feedback process is complex. Apart from one participant commenting that ‘feedback that aims at the regulation of the learning process’ was a missing option in question 17 for the kind of feedback provided, the remaining six comments participants clarify how they use feedback in their FO course. That is, when an assessment instrument is being graded, the kind of feedback and how that feedback is provided varies, for example, according to the type of assessment instrument, the ‘parts’ of a particular assessment instrument (i.e. some parts are automatically graded using CAA and others are graded by the instructor), the ‘number’ of students in the course and whether the feedback is provided after a question or after an entire assessment instrument has been completed.
Third, the remaining two comments do not appear directly related to the specific assessment schemes or feedback practices for their survey course context. One survey participant explained how FO mathematics instructors at their institution were soon ‘going to be allow[ed]’ to use invigilated assessment instruments, with the decision left to individual instructors. The other participant commented how the FO course context works well for ‘motivated students if the focus is on discussion rather than solving exercises’. They appeared to explain their view by going on to state that ‘answers to most problems’ can be solved online using Wolfram Alpha or other available mathematics software.

In summary, these comments reveal how many different facets exist in studying assessment practices. They present avenues for future research and suggest, by introducing complex factors affecting instructors’ decisions and actual practices, limitations to the thesis research.
Figure 3: Individual participants’ assessment practice by type of feedback used
Study I - R3: How are instructors using invigilation in their FO mathematics courses?

Previous studies investigated the proportion of FO courses not using any invigilation. Findings do not present any clear results (16% in Galante, 2002; 64% in Trenholm, 2007a). This study revisits this investigation and extends it by looking at which of the assessment instruments within a course are invigilated. Additionally of interest, it has been claimed that there is a greater emphasis on formative approaches to assessment practice when courses do not use any invigilation (Trenholm, 2007a).

Findings indicate that almost 40% (25/66 or 38%) of participants do not use any invigilation in their FO course. Making use of the feedback framework, the claim regarding a greater emphasis on formative approaches to assessment practice (Trenholm, 2007a) appears to be supported.

Instrument Perspective

In the first research question, where the use of invigilation is not considered, findings suggest most FO mathematics instructors emphasize summative-style instruments (i.e. final exam, tests and mid-term exam). For this question, when assessment instruments are differentiated based on the use of invigilation (see Table 17), some differences emerge. In particular, only a minority of participants using tests use invigilation (11/38 or 29%) and, on average, they weight their tests 20% more than those who do not use invigilation (57% compared to 36%). However, regarding those using final exams, while the majority of participants (33/45\(^{20}\) or 73%) use invigilation, some (12/45 or 27%) do not use any and, perhaps surprisingly, the average weighting assigned to invigilated final exams is not very different from those not invigilated (34% compared to 31% respectively). These differences are consistent with prior findings (Trenholm, 2007a) and suggest the use of invigilation presents a further level of complexity involved in investigating FO assessment practice.

\(^{20}\) Number discrepancy due to missing data.
### Table 17: Use of invigilation with instruments

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Not Using Invigilation</th>
<th>Using Invigilation</th>
<th>Missing Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>Number</td>
<td>Avg. Weighting (%)</td>
</tr>
<tr>
<td>Homework</td>
<td>100</td>
<td>50</td>
<td>27.7</td>
</tr>
<tr>
<td>Quizzes</td>
<td>100</td>
<td>32</td>
<td>20.2</td>
</tr>
<tr>
<td>Discussion</td>
<td>95</td>
<td>21</td>
<td>9.6</td>
</tr>
<tr>
<td>Individual Project</td>
<td>92</td>
<td>11</td>
<td>13.0</td>
</tr>
<tr>
<td>Other Assessment</td>
<td>75</td>
<td>6</td>
<td>14.8</td>
</tr>
<tr>
<td>Tests</td>
<td>71</td>
<td>27</td>
<td>36.1</td>
</tr>
<tr>
<td>Final Exam</td>
<td>27</td>
<td>12</td>
<td>31.4</td>
</tr>
<tr>
<td>Mid-Term Exam</td>
<td>27</td>
<td>4</td>
<td>17.2</td>
</tr>
</tbody>
</table>

### Individual Participant Assessment Practice Perspective

Results indicate that almost 40% (25/66 or 37.9%) of participants do not use any form of invigilation, contrasting with previous findings where about two-thirds (64%) did not use any form of invigilation (Trenholm, 2007a). Figure 4 displays participants’ assessment practices by instrument used. Figure 5 displays participants’ assessment practices by type of feedback use. The left side of each graph displays those participants using invigilation. The right side of the graph displays those participants not using any invigilation.

As shown in Figure 5, the ‘no invigilation’ group (right side) appears to make greater use of the richer type 2 feedback than those who do invigilate. To investigate this, tests are run comparing weightings associated with each of the three types of feedback used for those that use invigilation and those that do not. Results indicate statistically significant differences across two of the three types of feedback. First, on average, the assessment weighting associated with type 0 feedback for those not using any invigilation (Mdn=0.00) is found to be significantly different from those using invigilation (Mdn=30.00), U=341.50, Z= -2.334, p < 0.05, r= -0.29. Second, on average, the assessment weighting associated with type 2 feedback for those not using any invigilation (Mdn=36.00) is found to be significantly different from those using invigilation (Mdn=20.00), U=353.00, Z= -2.127, p < 0.05, r= -0.26. Both effect sizes represent a small to medium effect. For those not using invigilation, compared to those using invigilation, this indicates less emphasis on type 0 feedback and more emphasis on type 2 feedback. With the assumption that a greater emphasis on richer feedback implies a greater emphasis on formative
assessment practices (e.g. Yorke, 2003), these results appear to support earlier claims (Trenholm, 2007a) that those choosing not to invigilate are placing greater emphasis on formative approaches to assessment practice. This use of invigilation will be further explored in the eighth research question.
Figure 4: Use of invigilation related to individual participants’ assessment practices (instruments used)
Figure 5: Use of invigilation related to individual participants’ assessment practices (type of feedback used)
Study I - R4: What kind of professional development opportunities are FO mathematics instructors receiving for their courses?

As discussed in the literature review, it has been argued that disciplinary differences have been largely overlooked in the implementation of educational innovations (e.g. Neumann et al., 2002) and that, in particular, this has been an issue with regards to FO mathematics instruction (Smith, Torres-Ayala et al., 2008). This question seeks to investigate how FO mathematics instructors are currently engaging in professional development (hereafter termed ‘PD’).

Participants are asked a range of questions concerning the nature of PD they have available to them and how they are used. Table 18 summarizes these findings (n=66). First, three activities are considered, as shown, followed by a summary of how participants engage in these activities. Second, the remaining columns detail the participants’ use of funding and/or incentives related to PD in FO instruction followed by a summary of how the discipline of mathematics is specifically addressed in any PD (i.e. ‘Maths Focus’). In most cases the table provides both the number and percentage of participants availing themselves of each specific opportunity (i.e. ‘Yes’), then where an opportunity is available but not used (i.e. ‘Available, but no’), then not available at all.

Table 18: Participants’ PD Activities

<table>
<thead>
<tr>
<th>PD Engaged With</th>
<th>Activities</th>
<th>Average Number of Activities/Participant (out of a maximum 3 possible activities)</th>
<th>Funding/Incentives</th>
<th>Maths Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Internal Workshops</td>
<td>External Workshops</td>
<td>Sabbatical</td>
<td>Travel Funds</td>
</tr>
<tr>
<td>Yes</td>
<td>74.2% (49)</td>
<td>51.5% (34)</td>
<td>3% (2)</td>
<td>1.3</td>
</tr>
<tr>
<td>Available, but no</td>
<td>10.6% (7)</td>
<td>30.3% (20)</td>
<td>27.3% (18)</td>
<td>0.7</td>
</tr>
<tr>
<td>Not available</td>
<td>15.2% (10)</td>
<td>18.2% (12)</td>
<td>69.7% (46)</td>
<td>1</td>
</tr>
</tbody>
</table>

On average, participants have two of the three PD activities listed available to them. Of these activities, participants are more often engaging in them (1.3 activities per participant) than not (0.7 activities per participant). In particular, when internal workshops for FO instruction are available, almost all (49/56 = 88%) of participants acknowledged participating. However, fewer than half (31/66 = 47%) of all participants report that PD opportunities with a disciplinary focus on mathematics are available, with about one quarter (18/66 = 27%) reporting actual
participation. While it appears the nature of FO instruction, in general, is being addressed with available workshops, it also appears that the nature of mathematics instruction within that modality, as reported and argued in the literature review, is not being addressed by the majority of participant institutions or, when available, by all participants.

Considering time spent on PD activities, only 45 participants provide an answer. Of these 45, the average number of days of PD engaged in is 4.6 with a standard deviation of 8.4 showing wide variability.

This second part of the first study results provided background on the specific assessment instruments and practices used in a single FO course context for each participant. The next part uses this information with the ATI and S&B framework to analyze participants’ approaches to teaching and assessment in the same course context.

9.3 Part III: Approaches to Assessment Practice

This final part of the first study provides results as they pertain to research questions R5, R6, R7 and R8. These questions seek to investigate how FO mathematics instructors are approaching their assessment practice in the same single FO course context used in the second part. Where the feedback framework introduced in the previous part of this study provides some insight, this part extends the investigation by employing two additional analytic methods: the ATI and S&B frameworks, both detailed in the literature review and the methodology.

The limited literature on tertiary mathematics assessment practice, which primarily concerns F2F instruction, provides an idea of how participants may be approaching assessment in their FO courses. These findings indicate that tertiary mathematics instructors tend to use a transmissive (Lueddeke, 2003) and teacher-focused (Lindblom-Ylänne et al., 2006) approach to teaching whereby the application of course material and gaining factual knowledge are essential objectives of instruction (Barnes et al., 2001; Cashin & Downey, 1995). In particular, as recently confirmed in a large scale survey of US universities (Nelson Laird et al., 2008), tertiary mathematics instructors are found to use approaches to teaching that reflect an orientation to ‘knowledge reproduction’. As such, given participants’ prior F2F teaching experience, it may be expected that the present survey participants also orient their assessment practice according to these findings. This remaining part of the first study attempts to gain some insight into the nature of FO mathematics instructors’ approaches to teaching and assessment practice.
To conduct this investigation, survey data and results related to the two ATI and single S&B measures (here termed the ‘approach measures’) are presented and discussed. First (R5), for each of the six S&B questions, it is determined if the majority of participants’ responses are found either on what is considered to be the KR or KC half of the orientation continuum (see Table 19). Where possible, these results are related to claims and findings in the literature. Considering all six dimensions, overall results suggest that participants do not orient their assessment practice to KR, although it is unclear if they orient their practice to KC. Lastly, by means of stacked column charts, the relationship between participants’ S&B measures and their individual assessment practices are investigated. Second (R6), overall results for each ATI scale measure are presented. Third (R7), tests for correlational analysis are run to investigate the relationship of each of the ATI scale measures to the S&B measure. These results provide some support for the validity of the S&B measure. Additionally, tests for correlational analysis are run to investigate the relationship between the three feedback measures and each of the three approach measures. While the expectation is that the approach measures are related in some way to one or more of the feedback measures, no such association are found.

Finally (R8), all six study measures (i.e. the three feedback, two ATI and one S&B measure; here termed the ‘study measures’) are used to investigate specific assessment practices identified in the literature for their potential influence on the quality of student learning. Based on this investigation, with the caveat concerning the limitations outlined in the methodology, five significant relationships are discovered which largely appear to support some claims in the literature.

**Study I - R5:** *Using the findings of Samuelowicz and Bain’s (2002) ‘Identifying academics’ orientations to assessment practice’ study as a framework, how are FO mathematics instructors approaching assessment in their courses?*

**Individual Question/Dimension Perspective**

The following matrix (see Table 19) presents a novel display of overall approaches to assessment practices. It places the number of participants selecting specific responses within the original S&B framework that identifies six belief dimensions along six orientations to assessment practice (n=70; see Appendix B for the six S&B questions with possible responses in accordance with the original framework).
Table 19: S&B matrix summarizing participants’ responses with red line viewed as breakpoint separating majority KR from majority KC views

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Reproducing Bits of knowledge</td>
<td></td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>3</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>2: Reproducing structured knowledge</td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Applying structured knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4: Organizing subject knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5: Transforming discipline knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6: Transforming conceptions of the discipline/world</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to S&B, the first three belief dimensions concern the participants’ views of mathematical knowledge in the assessment context with the remainder referring to aspects of mathematics assessment practice in general. The results detailed below consider how the number of responses split either towards KR or KC (as delineated by the red line shown in Table 19). As will be shown, in contrast with expectations from F2F practice, findings do not suggest a similar overall orientation to KR.

Below, results are presented by first re-stating each question with possible responses in rank order from KR to KC (see Appendix G for survey questionnaire response order). Numbers after each response correspond to the number of participants making that particular selection and leaving a comment. Following this, for each dimension, the overall number of responses are discussed and where possible related to previous findings in the literature.
1. (Q10) In your approach to assessment, which of the following descriptions best describes how you view the nature and structure of mathematical knowledge?

a. It is external to students and is a collection of important definitions, concepts, techniques, methods and theories (26 with 5 comments).

b. It is external to students and is a coherent body of knowledge structured by experts in the field (6 with 1 comment).

c. It is something that is internalised, reorganised, and reconstituted in the process of learning (38 with 1 comment).

While results from the first question indicate the majority view the nature and structure of mathematical knowledge in assessment as requiring construction/transformation, the bipolar pattern suggests a more complex finding. That is, while most participants’ responses indicate a KC orientation, many participants’ responses indicate a KR orientation, and very few are found to choose the middle ground where expert-structured knowledge is emphasized. Given this is the only dimension with such a response pattern, this suggests that this is also the dimension with the least amount of consensus about how assessment is approached.

However, the majority of participants appear to emphasize mathematical knowledge, in the assessment context, as something that needs to be internalized (i.e. structured by the student). This appears to contrast with other findings that indicate students view mathematical knowledge as expert-structured (Crawford, Gordon, Nicholas, & Prosser, 1994; Fey, 1989; Lampert, 1990) suggesting that while participants in the FO context may be orienting themselves one way, students may be orienting themselves in the opposite direction.

Four of the five comments left by participants who selected the orientation associated with the KR extreme of the continuum (i.e. ‘It is external to students and is a collection of important definitions, concepts, techniques, methods and theories’) appeared to provide reasons for why these participants made this response selection. These reasons include responding in accordance with the type of assessment instrument (i.e. ‘exams’) they were considering, the way the instrument is delivered (i.e. use of CAA) and the nature of the specific course (i.e. ‘for this course”). These comments suggest the meaning of ‘approach’ may not be clear; that participants view ‘approach’ as attached to a specific context rather than an overarching orientation. For the fifth comment, the participant stated that ‘all of the above’ were applicable.
The remaining two comments, left by participants selecting the responses associated with the mid-range orientation (i.e. ‘It is external to students and is a coherent body of knowledge structured by experts in the field’) and the KC extreme (i.e. ‘It is something that is internalised, reorganised, and reconstituted in the process of learning’), both suggest these participants misunderstood the question by referring to their view of how students learn the nature & structure of mathematical knowledge rather than what it looks like. That is, for the participant selecting the mid-range, their comment suggests their response is a compromise between what they term their ‘goal’ of internalizing and reorganizing (i.e. KC) and their view of the nature & structure of mathematical knowledge as ‘due to the efforts of experts’(i.e. KR) and as a ‘collection of definitions’ and ‘concepts and techniques’. In the remaining comment, the participant again refers to process – simply stating: ‘I use Piaget’s model of reflective abstraction’ – when selecting the response associated with the KC extreme.

2. (Q11) In a typical assessment question, which of the following descriptions best describes how you assess your students?

a. Students draw on information presented in a single lecture, tutorial, practical session or chapter (21 with 2 comments).

b. Students draw on information presented in many sources, but all within the course/module (41 with 2 comments).

c. Students integrate information from many sources, from more than one subject, and/or from their own experience (8 with no comments).

Results from the second question focus on the degree of integration of knowledge expected in a typical assessment question. Overall findings indicate a majority of participants require their students to ‘draw on’ information coming from ‘many sources, but all within the course/module’ when answering a typical question. Looking at response numbers, there appears to be a majority of participants adopting a KR orientation with few (8) participants orienting their assessment approaches to transforming ‘discipline knowledge’ and ‘conceptions of the discipline/world’, while many (21) are found to be oriented to ‘reproducing bits of knowledge’. However, in balance, with most (41) in the wide mid-range, the actual split to KR or KC remains unclear.

Despite this uncertainty, when answering a typical assessment question, most participants appear to require their students to ‘draw on’ information that is, first, either provided within the course or, second, presented in a single instructional activity. Seldom do they require students to integrate information from many sources, from outside their subject and/or from their own
experience. To the extent that these selections reflect an information transmission/teacher-focused orientation (e.g. lecture notes are expected to be used as the basis for solving all problems), this is consistent with earlier findings regarding approaches to teaching in ‘hard’ disciplines (Lindblom-Ylänne et al., 2006).

Consistent with the first S&B question comments, participants’ comments left with this question suggest their responses would be different depending on the particular aspect of assessment practice the question is directed at (e.g. type of assessment instrument used). For the two participants selecting the response associated with the KR extreme (i.e. ‘Students draw on information presented in a single lecture, tutorial, practical session or chapter’), one participant commented that ‘applications come from the book’, perhaps differentiating their assessment questions as oriented to KR but non-assessment questions as not oriented to KR, however this is unclear. The other comment as well as the two associated with a mid-range orientation response selection (i.e. ‘Students draw on information presented in many sources, but all within the course/module’), similar to comments left with the previous S&B question, suggest participants’ responses reflect the context of assessment, which for these participants refers to the type of assessment instrument (i.e. quiz vs. exam vs. discussion) being emphasized in their course.

3. (Q12) In your approach to assessment, which of the following descriptions best describes how you assess your students?

a. I assess students on whether they can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes (17 with 4 comments).
b. Students apply well known techniques, methods, laws, principles, or explanations to unseen standard problems (46 with 5 comments).
c. Students apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems (7 with 1 comment).

Results from the third question focuses on the degree of transformation of knowledge. Findings indicate that the majority of participants require students to apply structured knowledge in their assessment approaches – implying the majority assess their students with a view towards the KR side of the continuum. Considering the number (63) that view their approach to assessment as focused upon assessing students’ ability to either reproduce what they have been presented in lectures or textbooks or apply procedures to different but ‘standard problems’, this appears consistent with other findings in the literature that have shown an emphasis, in mathematics, on content knowledge and application (Barnes et al., 2001).
In summary, findings for the first three S&B questions focused on views of mathematical knowledge in the assessment context show mixed results. The majority of participants view the nature and structure of mathematical knowledge (first question) with a KC orientation and the degree of transformation of knowledge (third question) with a KR orientation. However, regarding the degree of integration (second question), participants’ views are unclear.

This question, along with the question on the ‘use of feedback gained from assessment’, generated the highest number (10) of comments. Of the four left by participants who selected the response associated with the KR extreme (i.e. ‘I assess students on whether they can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes’), two of these comments suggest reasons for selecting their response – that the nature of the course ‘publisher’ (i.e. textbook) and ‘software product’ constrains the ability of participants to approach their assessment practice differently. For the other two comments, participants state, similar to comments left with other S&B questions, that the ‘level’ of the course and the dominant type of assessment instrument used influences whether open-ended problems can or may be asked. That is, because of these factors, open-ended problems are ‘rarely’ used or not the ‘majority’ of the questions asked.

The five comments left by participants who selected the response associated with the mid-range orientation (i.e. ‘Students apply well known techniques, methods, laws, principles, or explanations to unseen standard problems’) largely provide reasons why one of the other responses may also apply to these participants. That is, two participants’ comments suggest the response associated with the KR extreme is also applicable and, again, related to the ‘level’ of the course and type of assessment. The remaining three participants’ comments suggest the response associated with the KC extreme is also applicable, again, depending on the type of assessment instrument (i.e. for ‘projects’ or ‘online discussion’ where open-ended questions predominate).

Finally, for the remaining comment associated with the KC extreme (i.e. ‘Students apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems’), the participant gave reasons for selecting this response such as using ‘Minitab’ software to ‘draw conclusions’ and using a range of assessment instruments associated with this orientation (e.g. ‘essays’, ‘critical thinking questions’).
4. (Q13) In your approach to assessment, which of the following descriptions best describes how you view the difference between good and poor answers?

a. The difference between good and poor answers lies in the quantity of content correctly recalled (3 with no comments).
b. The difference lies in the accuracy and relevance of what is reproduced (25 with no comments).
c. Good answers are purposeful and justify the content used, whereas poor answers do neither of these things (42 with no comments).

While the previous three dimensions examine belief dimensions directly pertaining to mathematical knowledge, the remaining three dimensions examine further aspects of mathematics assessment with the present fourth S&B question focused on participants’ views regarding the differences between good and poor answers. Given few responses (3) are found in the most KR orientation (i.e. reproducing bits of knowledge), this suggests participants seldom view differences between good and poor answers simply by whether they can reproduce discrete or unstructured bits of information. However, overall findings are not clear given the third response, which straddles the line that breaks the orientations to either KR or KC, contains the majority of participants (42).

This was the only question where no comments were left. This may suggest, of the six S&B questions, this one was the most easily understood by participants.

5. (Q14) In your approach to assessment, which of the following descriptions best describes how you use assessment in teaching and learning?

a. Students have to be forced to study, and I use assessment as the best tool to achieve this (0).
b. Assessment forces students to study, and marks give them an indication of the progress made and reward their efforts (15 with 4 comments).
c. Assessment is an integral part of teaching and learning, a means of helping students learn (55 with 5 comments).

This fifth question shows the majority (55) of participants viewing their assessment practice as oriented to KC with no participants selecting the orientation most associated with KR (Note: this is also the only question with a response which no participant selected). These results show a relatively high level of consensus among participants regarding the role of assessment in teaching and learning. Few participants view assessment as a means to motivate students to
study or reward their efforts. Instead, the role of assessment is seen more as an important aspect of the teaching process and a means to further student learning. This seems to imply a prevailing view of assessment being used for learning (vs. of learning). This would appear to support claims regarding an emphasis on formative approaches to assessment in FO mathematics courses (Trenholm, 2007a).

This question generated the second highest number of comments (9). Four comments were left by participants selecting the response associated with the mid-range orientation (i.e. ‘Assessment forces students to study, and marks give them an indication of the progress made and reward their efforts’). In three of these, participants state that they would have selected more than one response, with two stating their other response selection would be the one most associated with the KC extreme. For the remaining comment the participant appeared to be attempting to clarify that although ‘assessment is external incentive to help students do what is necessary’, students eventually have ‘to want to learn’.

Of the five comments left by participants selecting the response associated with the KC extreme (i.e. ‘Assessment is an integral part of teaching and learning, a means of helping students learn’), four suggest these participants would also have chosen one of the other two responses associating assessment with ‘forcing’ students to study. That is, apart from the selected description of how they use assessment in teaching and learning, four participants also associate assessment with getting students, for example, ‘to do’ or ‘get involved’ in their learning, with two of these particularly associating this purpose with FO courses and two others associating iterative characteristics of their assessment practice with the process of learning. However, in seemingly stark contrast to the other four, one participant states: ‘Assessment measures what is going on; it does not force students to do anything’.

One further theme, which may be identified by two participants’ comments, highlights the tension between some ideal (or belief) vs. actual practice. One participant, selecting the response associated with the KC extreme, commented this selection reflected what s/he believes, not what they considered to be their actual practice. Similarly, another participant, this time selecting the response associated with the mid-range orientation, comments about what they would have selected in a seemingly ideal world vs. what they know of students in ‘day-to-day’ practice.
6. (Q15) In your approach to assessment, which of the following descriptions best describes how you use feedback gained from assessment?

a. I use feedback from student performance as a means of altering or adjusting my teaching approach (18 with 2 comments).

b. I use feedback from student performance to change my own or my students' actions (4).

c. I use feedback from student performance to monitor students' learning and to help them improve (42 with 8 comments).

d. I use feedback from student performance to challenge students' existing ideas and understandings (6 with no comments).

The sixth and final question – the only dimension with four possible responses – examines participants’ views regarding the use of feedback gained from assessment. Here, the majority (42) select the ‘transforming discipline knowledge’ orientation, indicating a majority of participants view the role of assessment and use of feedback with a KC orientation.

This question, along with the question on the ‘degree of transformation of knowledge’, generated the highest number (10) of comments. However, differentiated from all other S&B questions, these comments appeared to be the most succinct and consistent. That is, of the ten participants who left comments, seven used very similar succinct statements (e.g. ‘I do all the above’, ‘I actually do all 4’…) expressing that all four responses reflected their current practice. Of the remaining three participants’ comments, two stated that, along with choosing the most popular response associated with the KC extreme (i.e. ‘I use feedback from student performance to monitor students' learning and to help them improve’), they would also have chosen the response associated with the KR extreme (i.e. ‘I use feedback from student performance as a means of altering or adjusting my teaching approach’). While participants’ responses for this question suggest a somewhat clear consensus view, the nature of these comments suggest the S&B belief dimension (‘use of feedback gained from assessment’) represented by this question needs further study.

**Summary**

Considering results for all six questions, when approaching their assessment practice, the majority of participants view mathematical knowledge as something to be internalised, a typical assessment question as requiring students to 'draw on' information and student assessment focused on the ability to reproduce and/or apply ‘structured knowledge’. This is done so as to
produce answers differentiated by purposeful and justified use of content with the role of assessment, in particular, and feedback, to a large extent, regarded by participants as integral for student learning.

Overall comments suggest reasons why some participants may have been struggling to answer the S&B questions. These reasons include comments that suggest responses may be different depending on which assessment instrument(s) the response refers to or the ‘level’ of mathematics. In addition, responses reflect the constraints of the assessment delivery system (e.g. software or textbook) or the participants’ practice (but not their belief) or belief (but not their practice). In summary, many of these comments appear to suggest participants view ‘approach’ as attached to a specific context rather than an overarching orientation and, overall, that changes need to be made to the S&B questions if the questionnaire is to be used again in the mathematics assessment context.

In balance, the evidence does not suggest a clear orientation to KR (see Table 20) as expected from the literature (i.e. for F2F instruction, Nelson Laird et al., 2008). This raises questions about some possible influence, for example, from an emphasis on constructivist pedagogy in FO instruction (Anderson & Elloumi, 2008). This question is further explored in the second thesis study.

Table 20: Overall summary of participants' orientations

<table>
<thead>
<tr>
<th>Orientations to assessment practice</th>
<th>Belief Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>KC</td>
</tr>
</tbody>
</table>

*Individual Participants’ Assessment Practice Perspective:*

As discussed in the methodology, a single quantitative measure of participants’ approaches to assessment was created for use in the analysis. Overall participants’ S&B measures are found to range from 12 to 31.5 (out of a possible range of 7.5 to 31.5) with a mean of 21.45 and standard deviation of 4.22. A Shapiro-Wilk test for normality is run and the S&B measures, W(70) = 0.980, p > .05, are found not to be significantly different from a normal distribution.
For comparative purposes, individual assessment feedback practice is then related to S&B measures. Conceptually, there is an expectation that the S&B measure would be related, in some way, to one or more of the feedback measures. First, Figure 6 and Figure 7 are constructed to display participants' assessment practices associated with each instrument used and then by each kind of feedback used. These charts order participants from highest to lowest S&B measure (and then similarly by homework weighting) and, as seen, little or no pattern appears discernible. Second, tests for correlation are run to verify that there is indeed no association between each of the three feedback measures (i.e. weighting associated with type 0, 1 and 2 feedback) and the S&B measures. No significant correlations are found. This suggests that the kind of feedback used is not related to the approach to assessment as measured by the S&B measure.
Figure 6: Individual participants' assessment practice related to S&B measure
Figure 7: Individual participants’ use of feedback related to S&B measure
**Study I - R6: How are FO mathematics instructors approaching teaching in their courses as measured by Prosser and Trigwell’s (2004) Approaches to Teaching Inventory (ATI)?**

The ATI provides four subscale measures classified on two scales. The first scale measures how teaching approaches emphasize conceptual change with a student focus. The second scale measures how teaching approaches emphasize information transmission with a teacher focus. This section provides descriptive statistics on each of these scale measures (see Appendix J for full statistics related to all 16 questions that make up the full inventory).

As shown in Table 21, the participants’ CCSF measures, W(70) = 0.972, p > 0.05, and ITTF measures, W(70) = 0.979, p > 0.05, are both found not to be significantly different from a normal distribution. With the previous results from the normality test for the S&B measure, this means Pearson’s test for correlation may be used.

**Table 21: Overall scale measure descriptive statistics**

<table>
<thead>
<tr>
<th>ATI Scale Measures</th>
<th>Overall Measure (possible range of 8 to 40)</th>
<th>Test for Normality (Shapiro-Wilk)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D</td>
</tr>
<tr>
<td>Conceptual change/student-focused (CCSF)</td>
<td>26.0</td>
<td>5.9</td>
</tr>
<tr>
<td>Information transmission/teacher-focused (ITTF)</td>
<td>26.8</td>
<td>4.5</td>
</tr>
</tbody>
</table>

As with the S&B measure, a correlational analysis is run to investigate the relationship between each of the ATI scale measures and each of the three feedback measures. And, as with the S&B measure, no statistically significant correlations are found. This suggests that the kind of feedback used is also not related to the approach to teaching as measured by the ATI.

With no association found between the feedback measures and any of the approach measures, this suggests one of two things: that there is a problem with ‘feedback literacy’ (Havnes, Smith, Dysthe, & Ludvigsen, 2012, p.26) among participants or that there are problems with one or more of the frameworks or how they are operationalized.

Regarding the S&B framework, there is evidence suggesting both issues. First, in support of an issue of feedback literacy, a test of Cronbach’s alpha with item deletion is run and, of all six S&B questions, the reliability is found to improve the most if the S&B question on the use of feedback was deleted (i.e. from .434 to .515 compared to the next highest value, .482 for deleting the third question). This means, of all six S&B questions, participants’ responses
concerning their use of feedback appears least related to their overall approach to assessment, as quantified by the S&B measure. Second, in support of problems with the S&B framework, some evidence suggests this question is not fully addressing the nature of mathematics instruction. Specifically, this question receives one of the highest numbers of optional comments (see Appendix K) with the most consistent theme reflecting a lack of satisfaction with the response options – almost all comments (8 of 10) are responses such as ‘I do all of the above’. One possible issue, that this researcher discussed with Samuelowicz (personal communication, 2010), is that there is no specific orientation for using assessment just for monitoring. This may be prevalent in mathematics where there is an emphasis on summative assessment (i.e. the objective is to judge and not necessarily seek to improve student understanding). In general, issues with the use of the S&B framework in mathematic also appear demonstrated by comments provided with S&B responses. Despite these issues, however, given the responses for this question are found to strongly favour a single view and a lack of association is found with the ATI as an established psychometric instrument, the balance of evidence appears to suggest this is primarily an issue of feedback literacy. This is further explored in the second study.

Finally, in question 20, only three participants left comments related to the ATI inventory (see Appendix F). First, one participant stated question ‘o’ of the ATI (‘I feel that it is better for students in this subject to generate their own notes rather than always copy mine’) was inappropriate ‘when course materials are provided’, as may be the case with packaged textbook materials. Second, one participant appeared to be offering reflections, related to the ATI questions, clarifying their role and that of their students. For example, stating ‘I believe that an instructor...’ and then ‘I feel that students...’ Finally, one participant comment expressed interest in seeing the results of the survey. Apart from the comment about question ‘o’, the relatively small number of comments suggest most participants considered the ATI an appropriate instrument for measuring their approaches to teaching.

Study I - R7: How do findings in question six relate to findings in question five?

Results indicate statistically significant correlations between the S&B measure and the two ATI scale measures. In the following section these findings are outlined. First, Figure 8 and Figure 9 display the two scatter plots with associated lines of regression.

21 i.e. 42 participants selected “I use feedback from student performance to monitor students’ learning and to help them improve”
Figure 8: Scatter plot of CCSF vs. S&B measure with line of regression
With the assumption of normality for all three measures, two Pearson correlation tests are run to determine the degree to which the S&B measure is correlated to each of the two ATI measures. As a result significant correlations are found between the S&B measure and both the CCSF, $r = .391$, $p<.001$, and the ITTF, $r = -.436$, $p<.01$ measures.

Conceptually, the S&B measure is expected to be significantly and positively correlated with the conceptual change/student-focused scales and significantly and negatively correlated with the information transmission/teacher-focused scales. As shown, both correlation coefficients are in the expected directions and both statistically significant at the 0.01 level. With the ATI as an established psychometric instrument, these findings suggest the S&B measure has some validity and is measuring a similar underlying construct. That is, overall, for example, participants who tend to score higher on the ITTF measure tend to score lower on the S&B measure (i.e. towards KR) and those who tend to score higher on the CCSF measure tend to score higher on the S&B measure (i.e. towards KC).
Finally, Table 22 summarizes the correlational analyses. As shown, while significant relationships are found between the S&B and each of the two ATI measures, no significant relationship is found between any of the approach measures and any of the feedback measures.

Table 22: Correlation matrix of study measures

<table>
<thead>
<tr>
<th>Approach/Study Measures</th>
<th>Feedback Measures</th>
<th>ITTF</th>
<th>CCSF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>ITTF</td>
<td>no sig</td>
<td>no sig</td>
<td>no sig</td>
</tr>
<tr>
<td>CCSF</td>
<td>no sig</td>
<td>no sig</td>
<td>no sig</td>
</tr>
<tr>
<td>S&amp;B</td>
<td>no sig</td>
<td>no sig</td>
<td>-0.436 with p=.000</td>
</tr>
</tbody>
</table>

**Study I - R8:** *Are there any statistically significant differences in any of the study measures based on usage of invigilation, a greater variety of instruments, quizzes or discussion (the latter two as weighted instruments)? When used, is the weighting of either quizzes or discussion related to any of the study measures? Do these findings support prior claims and findings?*

In the literature review, the use of invigilation, a variety of assessment instruments, quizzes and discussion were identified for their potential impact on student learning. The use of invigilation has been discussed in the literature review with regards to an unknown influence on student learning (see results for R3, Trenholm, 2007a). With regards to a potentially positive influence on student learning, both the use of FO discussion (e.g. in general, see Swan, 2003; for mathematics, see Pomper, 2007) and the use of a variety of assessment instruments are widely discussed (e.g. Gikandi et al., 2011). Finally, with regards to a potentially adverse impact on student learning, the use of quizzes (e.g. Sangwin et al., 2010) is also discussed.

This section explores these factors in relation to the six study measures. In the first analysis, usage is investigated to see whether any of the study measures differ significantly based on whether invigilation, a variety of assessment instruments, quizzes or discussion are used. In the second analysis, the relationship between how quizzes and discussion are each emphasized and each of the study measures is investigated for any significant associations. Results confirm and conflict with claims and findings in the literature.
Use of Invigilation

The use of invigilation was originally covered by the third research question. There it was found that those not using any form of invigilation were significantly more likely to employ richer feedback (i.e. less type 0 and more type 2). Here the three remaining study measures (i.e. the S&B and two ATI measures) are investigated to see if there are any significant differences when a participant does or does not use invigilation.

Based on these tests, one significant difference is found. On average, the S&B measures of participants who did not use invigilation (M = 19.86, SE = 0.79) are lower than the S&B measures of participants who did use invigilation (M = 22.37, SE = 0.66). This difference is significant t(64) = -2.40, p < 0.05 and represents just under a medium-sized effect r = 0.29. If, as this difference suggests, instructors not using invigilation are significantly less oriented to KC, this presents some question about the increased use of hints and comments (i.e. type 2 feedback) and whether this feedback is actually linked with knowledge construction as expected (Butler & Winne, 1995). One possible explanation is that this feedback is not directed at student learning processes (feedback about the processing of a task, see Hattie & Timperley, 2007) but at validating students’ work and identities by means of establishing and maintaining contact that is more relational rather than academic. The use of invigilation is further explored in the second thesis study.

Use of a Variety of Assessment Instruments

The literature links the use of a variety of assessment instruments to effective assessment practices that produce deep learning (e.g. Gikandi et al., 2011). Related to this claim, tests are conducted to see whether the use of a greater variety of assessment instruments is associated with any of the study measures. Consistent with previous findings (Trenholm, 2007a), the number of different instruments used per participant range from one to seven, have a mean of 3.82 and a median of 4. Based on this information participants are split into two groups. The first, considered low to average, are those who use one to four instruments. The second, considered above average, are those that use five to seven instruments. Using these groupings, tests are run and one significant difference is found. On average, the S&B measures of participants using five to seven assessment instruments (M = 22.89, SE = 0.73) are higher than the S&B measures of participants using one to four assessment instruments (M = 20.63, SE = 0.68). This difference is significant t(64) = -2.11, p < 0.05 and represents a small-sized effect r = 0.25. While this does not imply that the use of a variety of instruments leads to deep learning,
it is consistent with prior claims that associate the use of a greater variety of assessment instruments with the nature of participants’ assessment practices, as gauged by the S&B measure.

**Use of Quizzes**

The literature links the use of quizzes to FO mathematics instruction. First, quizzes have been considered a central characteristic of FO instruction (Engelbrecht & Harding, 2005a; Greenberg & Williams, 2008). Second, some concern has been expressed related to the combined effects of immediate feedback (Sangwin et al., 2010) and repetitive attempts (Dubinsky, 1991), both practices associated with the use of quizzes (Butler et al., 2008).

As detailed in the previous results section on assessment specifics (see Part II), contrary to expectations from the literature, results indicate only a slight majority of participants use quizzes (35/66 or 53%). Of those, at least three-quarters (25/3322 or 75.8%) provide immediate feedback, supporting the association of immediate feedback with the use of quizzes (no data was collected regarding the use of repeat attempts; e.g. Griffin & Gudlaugsdottir, 2006). Based on these numbers, the usage and weighting of quizzes is investigated. However, quiz feedback timing is not investigated due to an insufficient sample size (i.e. n = 8 for those that do not use immediate feedback).

To investigate the usage and weighting of quizzes, two different statistical tests are conducted. First, tests are run to see if any of the study measures differ significantly based on whether or not quizzes are used as a weighted assessment instrument. Based on these tests, none of the study measures are found to differ significantly based on whether or not quizzes are used. Second, only for those participants using quizzes, tests are run to see if any of the study measures correlate with the quiz weightings. Based on these tests, the quiz weightings are found significantly and positively correlated to the information transmission/teacher-focused (i.e. ITTF) approach measure, \( \rho = .386, p < 0.05 \). The following scatter plot (see Figure 10) displays this relationship.

---

22 The difference (i.e. n=35 vs. 33) is due to two of the participants who failed to provide the timing of their quiz feedback.
These findings appear to be consistent with some of the concerns in the literature regarding the use of quizzes. While the proportion of participants (i.e. 53%) using quizzes appears contrary to expectations from the literature, of those that do, results support concerns that quizzes may somehow be associated with poorer quality learning outcomes (Trigwell et al., 1998). As indicated, the more participants emphasize quizzes, the more they are found to espouse an information transmission/teacher-focused approach. These results suggest the issue is not whether quizzes are used but how they are used. This issue is further explored in the second study.

**Use of Discussion**

As discussed in the literature review, there is a general emphasis on the use of discussion in FO courses as a critical component of instruction (e.g. White & Liccardi, 2007). While the use of discussion has been linked to ‘learning effectiveness’ (Swan, 2003), in the FO mathematics instruction context, findings present a conflicting picture – with the balance of findings
suggesting the use of discussion in FO mathematics courses does not work well (e.g. Glass & Sue, 2008). For these reasons the use of discussion as a weighted assessment instrument is investigated.

First, tests are run to see if any of the study measures differ significantly based on whether discussion is used as a weighted assessment instrument. Based on these tests, none of the study measures are found to differ significantly based on whether discussion is used as a weighted assessment instrument. Second, for those using discussion, a correlational analysis is run on discussion weightings across all study measures. Based on these tests, the use of discussion as a weighted assessment instrument is found to have no significant relationship with any of the study measures.

Overall results appear consistent with prior claims and findings. First, in support of findings that discussion does not work well in FO mathematics courses, less than 40% (26/66 or 39%) of participants are found to use discussion as a weighted assessment instrument. Second, a lack of association between the use and weighting of discussion and any of the study measures suggests that the use of discussion is not linked to the quality of teaching or assessment practice. The use of discussion is further explored in the second thesis study.

Table 23 outlines all findings for R8. In summary, while no significant associations are found in relation to the use of discussion, significant associations are found for the remaining three assessment practices under consideration. First, when quizzes are used, participants’ ITTF measures are found to be significantly and directly correlated with how much they are weighted in the overall assessment scheme. Second, when invigilation is not used, participants’ S&B measures and the total weighting of course instruments associated with type 0 feedback is found to be, on average, significantly lower while the total weighting of course instruments associated with type 2 is significantly higher. Third, when participants use more than four instruments as part of their overall course assessment scheme, their S&B measures are found to be, on average, significantly higher than those using four or less instruments. These results are limited due to the large number of tests performed and the increased probability of false positives. That is, at the 0.05 level with 36 statistical tests, one significant finding is expected to have arisen strictly by chance.
Table 23: Summary of analysis and findings for Study I - R8

<table>
<thead>
<tr>
<th>Population differences with usage and correlation with weighting</th>
<th>Approach Measures</th>
<th>Feedback Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S&amp;B</td>
<td>ATI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ITTF</td>
</tr>
<tr>
<td>Use of discussion (n=26 using, 40 not using)</td>
<td>no sig</td>
<td>no sig</td>
</tr>
<tr>
<td>Weighting of discussion (n=26)</td>
<td>no sig</td>
<td>no sig</td>
</tr>
<tr>
<td>Use of quizzes (n=35 using, 31 not using)</td>
<td>no sig</td>
<td>no sig</td>
</tr>
<tr>
<td>Weighting of quizzes (n=35)</td>
<td>no sig</td>
<td>(p=.386) with (p=.022)</td>
</tr>
<tr>
<td>Use of invigilation (n=41 using, 25 not using)</td>
<td>(p=.019)</td>
<td>no sig</td>
</tr>
<tr>
<td>Use of a variety of instruments (i.e. using 1-4 instruments compared with using 5-7; n=66)</td>
<td>(p=.039)</td>
<td>no sig</td>
</tr>
</tbody>
</table>

9.4 Summary of Research Findings

The first study used an online survey to investigate the assessment practices of current FO mathematics instructors where the majority (76%) of the survey participants taught in US institutions. The results of this investigation are summarized below for each research question.

**Study I - R1: What instruments are FO mathematics instructors currently using to assess their students? How are these weighted?**

Consistent with what is known about F2F mathematics courses (e.g. Iannone & Simpson, 2011), summative-style assessment instruments were also found to be emphasized in FO courses. There was considerable variation between individual survey participants’ assessment schemes. A minority (39%) of survey participants use discussion as a weighted assessment instruments, about half of the proportion found in prior studies (Trenholm, 2007a). About half (53%) of survey participants used quizzes as a weighted assessment instrument, fewer than expected. In addition, the least favoured assessment instruments were two types of group-oriented instruments (group projects and group work activities) as well as two types of individually-oriented instruments (journaling and portfolio work).
Study I - R2: How are instructors using feedback in their FO mathematics courses?

A novel feedback framework was constructed and employed to analyze the use of individual assessment instruments and survey participants’ assessment schemes. Descriptive statistics provided information such as, when used in an FO course, an average of 15 homework assignments were assigned with mean feedback timing of 2 days and an average of 1.1 final exams were assigned with mean feedback timing of 4.8 days. Feedback practices provided an indication of how instruments were being used for either primarily formative or primarily summative assessment purposes. For example, when using homework, survey participants provided all three forms of feedback. However, with an increasing number from the poorest to richest types of feedback, most provided the richest form of feedback. In contrast, when using final exams, the reverse was found: a decreasing number from the poorest to richest types of feedback with most providing the poorest form of feedback. Finally, overall, most participants were found to use a combination of two or more types of feedback in their assessment schemes with type 2 feedback (hints and comments) emphasized more than type 0 and type 1 feedback.

Findings also suggest problems with feedback literacy (Havnes et al., 2012) given no significant relationship was found between the three feedback measures and any of the three approach measures. This suggestion is supported by Cronbach’s alpha reliability coefficients that found, for each of the six S&B questions, that survey participants’ approaches to using feedback is least related to their overall approach to assessment.

Study I - R3: How are instructors using invigilation in their FO mathematics courses?

About two-thirds (62%) of survey participants use some form of invigilation in their FO mathematics courses, more than expected based on the literature (Trenholm, 2007a). About one quarter of all mid-term and final exams (27%), most tests (71%) and all homework and quizzes are not invigilated. On average, those not using invigilation used significantly more type 2 (hints and comments) and less type 0 (correct/incorrect) feedback than those who did use invigilation.

Study I - R4: What kind of professional development opportunities are FO mathematics instructors receiving for their courses?

Almost all (88%) survey participants take part in internal workshops for FO instruction. However, only a minority (27%) participated in professional development focused on FO
mathematics instruction. Overall, survey participants reported taking an average of 4.6 days of professional development.

**Study I - R5:** Using the findings of Samuelowicz and Bain’s (2002) ‘Identifying academics’ orientations to assessment practice’ study as a framework, how are FO mathematics instructors approaching assessment in their courses?

In contrast with expectations from the literature (for F2F instruction, e.g. Nelson Laird et al., 2008), findings do not suggest a clear orientation to KR.

**Study I - R6:** How are FO mathematics instructors approaching teaching in their courses as measured by Prosser and Trigwell’s (2004) Approaches to Teaching Inventory (ATI)?

Participants’ mean conceptual change/student-focused (CCSF) scale measure was found to be 26.0 (SD = 5.9). Participants’ mean information transmission/teacher-focused (ITTF) scale measure was found to be 26.8 (SD = 4.5).

**Study I - R7:** How do findings in question six relate to findings in question five?

Both ATI measures (i.e. ITTF and CCSF) were significantly correlated, p<0.01 to the S&B measure. Both correlations were in the expected directions (ITTF: r = -.436; CCSF: r = .391). With the ATI as an established psychometric instrument, these findings suggest the S&B measure has some validity and is measuring a similar underlying construct. That is, overall, for example, participants who tend to score higher on the ITTF measure tend to score lower on the S&B measure (i.e. towards KR) and those who tend to score higher on the CCSF measure tend to score higher on the S&B measure (i.e. towards KC).

**Study I - R8:** Are there any statistically significant differences in any of the study measures based on usage of invigilation, a greater variety of instruments, quizzes or discussion (the latter two as weighted instruments)? When used, is the weighting of either quizzes or discussion related to any of the study measures? Do these findings support prior claims and findings?

On average, S&B measures of survey participants who did not use invigilation were lower than the S&B measures of survey participants who did use invigilation. On the surface this appears to contradict earlier findings where a greater use of richer feedback may be expected to be associated with a greater orientation to knowledge construction (e.g. Hattie & Timperley, 2007). One possible explanation is that this feedback is not directed at student learning processes.
(feedback about the processing of a task) but at validating students’ work and identities by means of constant contact.

Consistent with expectations (e.g. Harlen & James, 1997), on average, the S&B measures of survey participants using five to seven assessment instruments were higher than the S&B measures of survey participants using one to four assessment instruments.

Quiz weightings were found to be significantly and positively correlated to ITTF approach measures, $\rho=.386, p < 0.05$. Conceptually, with almost all quizzes being CAA-based, this is consistent with expectations (e.g. Paterson, 2002) that online quizzes lead to lower level learning and therefore a greater emphasis on these quizzes is associated with approaches more oriented to information transmission/teacher-focused teaching.

No significant relationship was found between use or weighting of discussion and any of the study measures.

9.5 Discussion – Further Research

From the first study several questions emerge regarding specific assessment practices. These questions are related to the following findings: First, findings suggested a decline in the use of discussion among SUNY instructors and, contrasting with expectations from the literature, no relationships were found between the use of discussion and any of the study measures. Second, in contrast with expectations from the literature, quizzes were found to be used by only about half of all participants. Additionally, an emphasis on quizzes was found to be correlated with the ITTF measure. Third, on average, compared to those that use invigilation, those not using any invigilation were found to be providing significantly more comments and hints as feedback, as measured by the feedback measures, but have significantly lower S&B measures. Fourth, findings suggest a possible issue of feedback literacy given no relationship was found between any of the approach and feedback measures. Finally, in contrast with expectations from the literature, participants S&B question responses suggest no clear overall orientation to KR. The second study will use interviews with a sample of survey participants teaching in US institutions to try to explain these findings and gain more insight into FO mathematics instructors’ teaching and assessment practices.
10. Study II Method and Methodology

10.1 Introduction

The first study investigated the specific ways instructors were using assessment instruments in their FO courses, the way these instructors were approaching their teaching and assessment practices and whether there were any significant associations between the different approach measures and particular assessment practices. This second study attempts to understand and explain some of the first study findings. As discussed in chapter seven, this combination of a quantitative followed by a qualitative study is considered an effective means of providing a greater understanding of the nature of current FO mathematics courses than would otherwise be achieved by either study alone.

To do this, an interview questionnaire is constructed, interviews are conducted with six US participants, and the responses are analyzed. From a US perspective, the findings provide some insight into the affordances and constraints of the current FO course context and consequently the nature of current FO mathematics teaching and assessment practices.

The following chapter details the interview methodology. First, the overall research design is detailed and the research questions are considered. Second, the construction of interview questions is discussed and related back to both the literature and the first study results. Third, the pilot interview process is described together with changes that were made based on the pilot. Fourth, the actual interview process is detailed including how ethical issues were considered. Fifth, information is provided about how the analysis was conducted and findings arrived at. The chapter concludes with a discussion on issues of validity and reliability.

10.2 Overall Interview Study Research Design

Assuming all participants are teaching or have taught F2F courses, the second study proposes (i) that instructors must modify their F2F assessment practice in order to adapt to the FO instructional context, and (ii) that this adaptation reveals constraints and possibilities that provide a window into the current nature of FO instructors’ assessment practice.

Based on this proposition, the second study seeks to explain the associations found in the first study as well as the nature of FO instructors’ current orientations to assessment practice. The second study focuses on the use of discussion, quizzes, invigilation and feedback as well as the
emphasis on KR and/or KC. In the next section the specific study propositions and associated research questions are detailed.

Directed by the first study findings an interview questionnaire is constructed with questions directed at asking participants how specific aspects of their assessment practices differ when they teach the same course in both contexts. In this way, participants’ F2F experiences provide a baseline through which their FO assessment practices can be described, reflected upon and analyzed. The interview questionnaire consists of nine questions (see Appendix L). Participants are separately asked all nine questions in the same order but in a semi-structured format which allows them to discuss possibly divergent issues as they arise.

As detailed in section 10.6, the analysis is conducted for each interview question (not each participant). Transcripts are grouped and analyzed using the constant comparative method (Boeije, 2002) to build explanations regarding the nature of current FO assessment practice. Emergent themes are of primary interest; however, where possible, other minority rival views are also considered.

10.3 Specific Theoretical Propositions and Research Questions

The following section details each of the five research questions addressed in the second study. Preceding each question a brief summary of the relevant background literature is given. This is followed by the specific study theoretical proposition(s), largely provided by means of the related first study finding(s). In all, five broad question areas (use of discussion, use of invigilation, use of quizzes, use of feedback and orientation to KC) were initially developed for use in pilot interviews.

10.3.1 Use of Discussion

*Literature Background:* In the FO instructional context, interactions between students and with the instructor tend to be emphasized (e.g. Swan, 2003). These interactions primarily take place in the form of discussions or ‘threaded discussions’. In mathematics, problems with communication make these interactions difficult as, for example, almost all communication is largely text-based and mathematical notation is often more difficult to communicate in the FO than in the F2F context. In the FO context, when pure disciplines (e.g. mathematics) are taught, they have been found to be less community-oriented than other disciplines (Smith, Heindel et
al., 2008). The balance of findings suggests discussion in FO mathematics courses is not working well.

**Study I Background/Study II Theoretical Proposition(s):** Findings indicate that a minority of participants use ‘discussion’ as a weighted assessment instrument. There appears to be no significant relationship between the use of ‘discussion’ as a weighted assessment instrument and any of the approach measures.

**Study II - R1:** How and why is discussion/interaction used?

10.3.2 Use of Invigilation

**Literature Background:** In increasingly market-oriented higher education practice, the use of invigilation runs counter to some institutional priorities (i.e. providing flexibility and convenience; Trenholm, 2007b).

Prior study findings indicate that many FO mathematics courses do not use any form of invigilation. FO mathematics courses are particularly susceptible to threats from cheating (Trenholm, 2007b). Compared to when invigilation is used, when no invigilation is used, summative-style instruments are used less often and, when used, weighted less (Trenholm, 2007a). This suggests a greater emphasis on formative-style instruments.

**Study I Background/Study II Theoretical Proposition(s):** Findings indicate that a minority of participants do not use invigilation. Those not using any form of invigilation are significantly more likely to use richer feedback – as defined by the study framework for feedback – and significantly less oriented to KC than those that do use invigilation.

**Study II - R2:** How and why are participants choosing to use invigilation?

10.3.3 Use of Quizzes

**Literature Background:** FO mathematics assessment practice is said to emphasize quizzes with immediate feedback considered as the ‘most basic assessment activity’ (e.g. Engelbrecht & Harding, 2005a, p.247).

**Study I Background/Study II Theoretical Proposition(s):** Findings indicate that just over a half of participants use quizzes as a weighted assessment instrument. For these participants the
weighting allocated to quizzes is significantly and positively related to the degree to which an information transmission/teacher-focused approach to teaching is used.

**Study II - R3: How and why are quizzes being used?**

### 10.3.4 Use of Feedback

**Literature Background:** In the FO context, there are at least two reasons why good feedback practice is prioritized. First, it is seen as a means to enable knowledge construction (i.e. ‘assessment for learning’, Gaytan & McEwen, 2007). Second, good feedback practice is considered vital for overcoming inherent challenges in the FO instructional environment where students are separated from each other and the instructor (Wolsey, 2008 as cited in Gikandi et al., 2011).

In the FO literature, there is a general emphasis on the importance of immediate feedback linked to learning effectiveness (Swan, 2003). However, in the CAA literature, immediate feedback has been found to be effective only for ‘lower-level’ learning tasks (Morrison, Ross, Gopalakrishnan, & Casey, 1995) whereas delayed feedback is seen to be effective for ‘higher-order’ learning (e.g. in general, see Kluger & DeNisi, 1996; in mathematics, see Simmons & Cope, 1993).

**Study I Background/Study II Theoretical Proposition(s):** Findings suggest there is no significant relationship between the kind of feedback instructors’ use and their approaches to teaching and assessment.

**Study II - R4: How and why is feedback being provided?**

### 10.3.5 Orientation to Knowledge Reproduction and/or Knowledge Construction

**Literature Background:** The nature of FO instruction emphasizes the role of students in the learning process (and de-emphasizes the instructor’s role). There is an expectation that FO students will be autonomous, self-directed and active in constructing their own knowledge (Anderson & Elloumi, 2008). For current tertiary mathematics instruction this is a significant shift for at least three reasons. First, information transmission and instructor-focused approaches appear to prevail in current tertiary mathematics instruction (Cashin & Downey, 1995; Lindblom-Ylänne et al., 2006). Second, it is has been found that constructivist approaches are used infrequently in tertiary mathematics instruction (Walczyk & Ramsey,
2003). Third, the influence of constructivism on mathematics education has been considered to be ‘waning’ (Confrey & Kazak, 2006).

**Study I Background/Study II Theoretical Proposition(s):** Overall, findings suggest no clear tendency to KR in participants’ FO assessment practices – as analyzed using the S&B framework.

**Study II - R5: How are participants’ approaching their FO course assessment practice?**

Based on all of these questions an initial questionnaire was constructed consisting of 13 questions. This number was reduced following an iterative review process which focused on ensuring that the interview questions were limited to addressing the research questions. While most questions remained the same, one major decision was made to break the question on feedback into three parts (i.e. ‘kind’, ‘purpose’ and ‘timing’). This was done so as to better address facets, identified in the literature, of how and why feedback is provided. As a result, a questionnaire with seven questions was finalized for use in pilot interviews.

10.4 Pilot Interviews

After an interview protocol was decided upon, two pilot interviews were conducted using the initial interview questionnaire. This section details this process along with subsequent changes that led to the development of the final interview questionnaire and protocol.

10.4.1 Pilot Interview Participants and Procedure

The first pilot interview was run with a local UK university lecturer, equivalent to an instructor at a 4 yr university, with experience in qualitative research in mathematics education, but who only taught F2F courses. This was followed, about one week later, with a second pilot interview with a US community college instructor who taught a hard science Geology course in both the F2F and FO modalities.

Pilot interviews were conducted using Skype, were semi-structured and the length of time it took to complete all the interview questions was recorded. Given the first pilot participant’s expertise in qualitative research, following his/her interview, s/he was invited to identify any possible areas of confusion and provide any advice or comments concerning the interview questionnaire and process. In the second pilot interview, no such feedback was solicited. However, in both
instances copious notes were taken. Based on these notes, as well as the feedback from the first pilot, the questionnaire underwent additional changes as detailed below.

10.4.2 Pilot Interview Feedback and Changes Made to Final Interview Questions

While the overall interview protocol (see Appendix L) was deemed satisfactory, the results of the pilot interviews suggested four significant changes to the interview questions. First, the language used for the question on KC was considered to be too technical and thus unsatisfactory for communicating the variation in the S&B scale (from KR to KC). As a result the language was changed so as to make the question understandable to someone not versed in education research. This led, for example, to dropping the term ‘knowledge reproduction’ and instead using the explanation for this term found in the S&B study and contextualized to mathematics (i.e. ‘reproducing important mathematics facts, procedures and skills’) – although care was taken not to diverge from the original meaning in the S&B framework. Second, based on the second pilot interview, the number of questions directed at the instrument level was found to be unrealistic. For example, the second pilot participant used seven different assessment instruments in his/her F2F course whereas two of the feedback questions asked for an instrument by instrument breakdown of feedback practice in both the F2F and FO contexts. This led to modifying these questions so that they addressed any differences directly rather than indirectly through a systematic review of each individual instrument in each context. Third, related to this number and the expected variety of assessment instruments, a table was incorporated into the beginning of the questionnaire whereby participants would be asked, at the onset of the interview, to provide their course assessment schemes for both the F2F and FO contexts. This also helped to frame the seventh question which sought instrument-by-instrument participant responses on approaches to assessment. In particular, for each instrument used, rather than seeking a narrative response, participants were simply asked if they viewed that instrument as more oriented to KR, KC or both equally. Finally, two additional questions were added at the end of the questionnaire to seek out overall participant views concerning what they considered to be the role of assessment in the FO context and what they considered to be some of the problems and benefits of assessment in the FO context. These questions were added to provide an opportunity for participants to express any overarching views. As a result of these changes the final interview questionnaire had nine questions (see Appendix L).
10.5 Interview Methodology

This section describes how and with whom the actual interviews were conducted. First, the process to ensure the interviews were conducted in an ethical manner is detailed. Second, the initial participant selection procedure is described. Third, the procedure for inviting participation is detailed. Fourth, characteristics of the final participants and their associated courses are provided. Fifth, the final interview protocol is described.

10.5.1 Ethical Issues

As with the first study, similar steps were undertaken to ensure that the research was conducted in an ethical fashion. First, university ethical guidelines for research were followed and an Ethical Clearance Checklist was completed and lodged in the department. Second, on both the email invitation and before the interviews began, participants were fully informed of the purpose of the study and how the collected data was to be used. This included being informed that the interviews were recorded, that participants could stop the interview at any time and that all collected data would be used anonymously and confidentially.

10.5.2 Initial Participant Selection Procedure

Interview participants originated from a list of 14 survey participants who voluntarily left their email addresses as part of the first study. From this list, a representative selection of candidates was sought. First, the 14 participants, together with their associated survey results and the calculated study I approach measures, were ordered according to their S&B measure. Second, because of this researcher’s familiarity with the US context, a decision was made to focus on the US participants. There were 10 US participants who were fairly evenly spread along the S&B measures. Third, for interview studies, Guest, Bunce, & Johnson’s (2006) experimental findings suggest that a careful and purposeful sample of six participants may be ‘sufficient to enable development of meaningful [‘high-level’ and ‘overarching’] themes and useful interpretations’ (p.78). Therefore, from these 10 participants, six candidates were sought representing both two- and four-year institutions, those who used and did not use invigilation\(^2\) and a range of approaches to assessment, as per their first study S&B measure. Limitations to this process will be discussed at the end of this chapter.

\(^2\) Expected use of invigilation is based on the first study survey course context.
10.5.3 Procedure for Initial Contact

These six potential participants were sent an initial email invitation. This email had three main purposes: to invite instructors to participate, to provide basic information about the interview (e.g. time commitment) and to satisfy ethical guidelines by informing participants how the interviews would be conducted and used. Lastly, it was made clear that participants were free to withdraw at any time during the interview.

Responses were received fairly soon after the initial emails were sent out with five of the six agreeing to be interviewed. An additional replacement candidate was then selected using the same criteria and subsequently agreed to be interviewed.

10.5.4 Interview Protocol

Interviews took place within a few weeks of the initial email response: potential participants were emailed at the end of December 2011 and all interviews were completed by the end of January 2012.

As with the pilot, all interviews were conducted using Skype\textsuperscript{24} and recorded on two separate recording devices\textsuperscript{25}. Participants were contacted at a pre-arranged time. Each interview began with an informal greeting and introduction to make participants comfortable and at ease, which was then followed by a standard introduction (see Appendix L). This included a reminder of ethical guidelines and an opportunity to ask, if need be, any questions concerning the interview. Next, a single course context was chosen for the interview\textsuperscript{26} with the objective of using a course that participants taught in both the F2F and FO contexts. Upon selecting the context, participants were then asked to outline their assessment schemes including instrument usage and weighting for both\textsuperscript{27} F2F and FO courses.

\textsuperscript{24} All interviews except one were audio only. One participant initiated audio with video and this was reciprocated.
\textsuperscript{25} By accident one of the interviews was not recorded. After consulting with a local expert on qualitative research, as the interview participant was willing, a second interview was conducted about one week later. Answers from this interview were compared with the notes and memory recall from the first interview and considered to be the same or similar to the first interview answers.
\textsuperscript{26} Research (e.g. course offerings listed on the institutional website) was conducted beforehand to determine which course(s) these participants taught in both modalities. This information was discussed with participants at the beginning of the interview and an appropriate course context was decided upon. In keeping with the interest in the first study, where possible, Calculus and Statistics courses were sought.
\textsuperscript{27} As previously discussed, one participant taught only one FO course which they also did not teach in the F2F context.
Once this information was gained, each of the questions was posed in a semi-structured format. Ample time was provided for participants to answer and, where deemed necessary, the interviewer probed for greater depth and clarity. In particular, with the question on KC (the seventh question), care was taken to ensure that participants understood that the framework represented a continuum and not an either/or choice. To do this, each participant was asked to confirm that they understood the framework as it was explained. After the last (ninth) question, the interview ended with thanks and, as many expressed an interest, a promise to share interview results after the thesis had been submitted. Finally, the audio was stored and backed up in preparation for transcription and analysis.

10.6 Data Analysis

Transcripts of all interviews were made and the analysis was performed on responses for the first seven interview questions. To prepare for this analysis, each of the transcripts was read to check whether answers to specific questions were elaborated upon or clarified while other questions were being answered. Where this was the case, these answers were copied and pasted to where the actual question was covered. For example, responses from the last two questions (i.e. questions eight and nine) were added to the other questions where these responses address or clarify earlier question responses. After this process was complete, the data was then organized into a matrix with six columns representing each of the first six interview questions and six rows representing each of the participants. From there each column of collected responses was imported into six different Atlas Ti files for analysis. The first six questions were analyzed using the constant comparative method (Boeije, 2002) to build explanations regarding the nature of current FO assessment practice. The seventh question was analyzed based mostly on a quantitative comparison of the identified approaches for each of the assessment instruments used in the F2F and FO courses.

To begin, using Atlas-Ti, the transcripts of participants’ responses for each of the first six questions were separately read and reread. After the second reading, as themes emerged, they were coded. For example, Figure 11 displays some excerpts, with initial coding, for the first question on the use of discussion.
Figure 11: Screenshot showing Atlas-Ti coded excerpt of interview responses for the first question on use of discussion (Note: A4 and B4 were the original codes used for participants P1 and P2, respectively)
Following this initial process of coding, additional readings of the transcripts for each of the six interview questions were used to verify and, where necessary, introduce additional codes or recode previous excerpts. For the question on the use of discussion, for example, this resulted in a total of 48 codes summarizing the ideas communicated by all six participants. Following this, these codes were further analyzed using the Atlas-Ti code family manager. This process was used to group codes into emergent themes. As shown in Figure 12, for the question on the use of discussion, 12 themes emerged out of the 48 codes.

When this process was completed for each of the first six questions, the analysis then moved to paper-based work. For each question this involved writing the themes on individual post-its (i.e. small pieces of paper with adhesive backing) and placing them on separate sheets of A4 paper. This allowed moving themes next to each other and facilitated a process of reflection regarding potential relationships between different themes.

However, before beginning the write-up, a decision was made concerning how to structure the results for each question. Using the process described in the previous paragraph and by reflecting on the research questions, three broad categories surfaced for structuring the write-up\(^28\): factors related to the use of the pedagogical practice, differences in that practice in F2F and FO courses and responses relating that practice to the quality of learning.

Using these three categories the emergent themes were then used to construct an initial narrative account of findings for each of the first six questions. This process of writing allowed for further reflection which led to re-writing and, where necessary, revisiting and changing family codes or the order in which themes were presented. For each question, this iterative process continued until the narrative was judged to be an accurate representation of participants’ claims and experiences. The end product was seven narrative accounts of findings where the seventh presents what was largely a quantitative analysis of participants’ responses.

\(^{28}\) With one exception: For structuring the results on participants’ use of quizzes, the use of CAA emerged as an additional category.
Figure 12: Screenshot of Atlas-Ti code family manager for the question on the use of discussion. The code families are shown in the upper left-hand corner. (Note: ‘Size’ and ‘Quotes’ refers to the number of codes and quotes, respectively, in a particular family). The lower right-hand corner displays a scrollable table of all codes found for the use of discussion.
10.7 Validity and Reliability

Though some researchers (e.g. Schwandt, 1990) argue that criteria for good qualitative work cannot be specified, others argue that standards of judgement do exist (e.g. Howe and Eisenhart, 1990). Miles and Huberman (1994), for example, provide a framework which considers five main issues for evaluating the ‘trustworthiness’ of qualitative research. This framework will now be used to discuss the validity and reliability of the second study.

The first issue, objectivity, concerns researcher bias. Two matters related to this issue were considered pertinent. Firstly, there was recognition that this researcher’s views of FO mathematics courses were informed by personal experience of teaching and conducting research in this area. In this respect apart from maintaining a level of personal self-awareness of any bias related to this experience, an effort was made at keeping an open mind to all possible outcomes. This included taking on board and acting on any comments provided by those supervising this research. Secondly, if the study was to be rerun or the data reanalyzed, the data was safely and securely retained and the method of data collection and analysis was detailed. This included taking hand-written field notes using a paper-based interview questionnaire for each interview participant, recording and carefully transcribing all interviews, and then backing-up all this data.

The second issue, reliability, concerns the level of consistency and care exercised in the study process. This issue was addressed in at least three ways. Firstly, as detailed earlier in this chapter, the second study interview questions were carefully linked to the second study research questions, and these research questions were linked to the first study findings as well as the literature. Secondly, using the list of those first study participants volunteering for a follow up interview (n=14), the second study interview participants (n=6) were purposefully selected using information about their background and assessment practice collected in the first study. Specifically, only US participants from public two and four year HE institution and a range of approaches to assessment and uses of invigilation (i.e. uses or does not use invigilation) were selected. The latter two characteristics were determined using participants’ first study S&B measures and information they provided on how they used invigilation. Finally, when the analysis was conducted the coding was repeatedly reviewed and, if necessary, changed or new coding was created. This was done at the onset of using Atlas-Ti to code the transcripts and identify families of codes. Here, changes in and/or the introduction of new coding mainly appeared to occur when newly introduced codes (or families of codes) called into question those previously introduced. Following this, further changes were made when the passage of time and
the process of writing up the findings in narrative format permitted further and deeper analysis and synthesis. However, while those supervising this research were actively involved in reviewing the write-up of the analysis, one concern about reliability is that no other observers participated in the initial coding process. The systematic nature of analysis, described above, was considered to at least partially mitigate for this concern.

The third issue, internal validity, concerns the truthfulness of the narrative findings. That is, do the findings present an explanation that makes sense to those studied as well as those reading the study? At least five actions were taken to address this issue. Firstly, interview participants were given an opportunity, with no email responses received, to challenge the findings when a summary was sent out to each participant in December 2012. Secondly, an effort was made to make explicit the background of the interview participants including their institutional, class and course contexts. This included detailing the background data collected and measures found in the first survey study. Thirdly, any limitations, such as participants’ use of CAA which will be covered in the results, were discussed. Fourthly, two additional interview questions (eight and nine) were included to provide an opportunity for interview participants to fully explain and/or correct their stated responses to the first seven questions. Finally, as will be covered in the discussion in Chapter 12, findings were related to the literature and current theory.

The fourth issue, external validity, concerns the generalizability of the findings or how they can be interpreted. The threat to external validity posed by the nature of sampling was considered to be the most serious issue related to the overall validity of the second study. While a detailed description of the interview participants including their approach measures and other data from the first study provided some ability to make a purposefully diverse selection of participants. And a sample of six participants was considered to be a sufficiently large enough sample size for providing ‘high-level, overarching’ themes and ‘useful interpretations’ (p.78) that help answer the research questions (Guest, Bunce, & Johnson, 2006). However, despite such measures, it was still not clear how this sample reflects the general population. To counteract this threat, three actions were taken. Firstly, as previously mentioned, to aid in making comparisons with other samples, the study process from data collection to analysis, as well as the nature of the participants and the context, was detailed. Secondly, findings were discussed in relation to the literature and prior theory and any links or conflicts were discussed. Thirdly, perhaps most importantly, a mixed methods research design was considered to counteract (Harris and Brown, 2010) this threat particularly, for example, when the findings from the first quantitative study
were used to direct the construction of the second study research questions and the selection of the participants.

Finally, the fifth issue, utilization, concerns the actual usefulness of the findings both to researchers and teaching practitioners. The actual usefulness largely remains an open question until findings are disseminated. However, initial face value usefulness was considered to be achieved because findings were intended, particularly with respect to mathematics assessment practice, to help explicate the affordances and constraints of the FO course context and suggest specific actions to aid in current practice.

In summary, findings need to be interpreted with respect to these limitations and the actions taken to address them, as well as the background context and characteristics of the interviewees.

10.8 Summary

- The overall interview design was introduced where it was proposed that investigating differences between F2F and FO assessment practices provided a means of explaining the nature of current FO assessment practice.

- As directed by the first study findings, the aspects of current FO assessment practice to be focused on in the second study were presented. That is, five specific theoretical propositions were presented together with five associated research questions.

- Based on these questions, and following pilot interviews, a nine-item interview protocol was designed where items largely focused respectively on the use of discussion, quizzes, invigilation and feedback as well as the emphasis on KR and/or KC.

- Six US participants, selected from varied backgrounds, agreed to participate in interviews which were conducted using Skype.

- Qualitative, and some quantitative, analysis was conducted. On the first six questions (i.e. investigating the use of discussion, quizzes, invigilation and feedback) the largely qualitative analysis produced six narrative accounts of findings. On the seventh question the largely quantitative analysis was presented in a narrative format.

- The chapter ends with a discussion on issues of validity and reliability, along with measures taken to address them.
11. Study II Results

Interviews were conducted with six US survey participants from the first study – ranging from approximately 42 to 93 minutes in length or just over 60 minutes on average – according to the protocol described in the previous chapter. This chapter details results of the analysis for each of the first six interview questions (see Appendix L; corresponding to the first four second study research questions in Table 24). Results for the seventh interview question (corresponding to the fifth research question) on overall approaches to assessment – where the analysis is less qualitative and more quantitative – are covered in chapter twelve which discusses the second study results.

Before presenting these results, the first two sections of this chapter provide some background information that helps contextualize the research and explain some of the terminology used by participants. Following this, in the next six sections, the analysis is presented. Three sections cover the use of discussion, quizzes and invigilation. The next three sections cover the use of feedback – the kind (and process), the purpose and the timing of feedback. Each of these six sections is structured similarly according to emergent themes. First, factors that participants identify as influencing their assessment practice in either context are detailed. Second, related participant-identified purposes are outlined. Third, related participant-identified differences between their F2F and FO courses are detailed (hereafter simply referred to as ‘differences’). Fourth, participants’ views concerning the overall quality of learning in FO courses is discussed and conclusions are drawn. Overall, by every practice investigated, participants describe the FO course context as a challenging environment for effective assessment practice. Moreover, compared to the F2F course context, findings suggest the FO course context presents a much more challenging environment for realizing deeper quality learning in mathematics.

29 There are four exceptions: First, no factors are identified or discussed related to the use of discussion. Second, the results on the use of quizzes also include a section on evidence for participants’ dependence on the use of CAA. Third, for the sections on the kind (and process) as well as timing of feedback, participant-identified purposes for using feedback are covered in the section on the purpose of feedback. Fourth, the sections on the kind (process) and timing of feedback have additional sections detailing, respectively, the most effective kinds of feedback and benefits of immediate feedback.
Table 24: Study II research questions and corresponding interview questions and analysis

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Interview Question(s)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1. How and why is discussion/interaction used?</td>
<td>1</td>
<td>Using constant comparative methodology</td>
</tr>
<tr>
<td>R2. How and why are quizzes being used?</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>R3. How and why are participants choosing to use invigilation?</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>R4. How and why is feedback being provided?</td>
<td>4-6</td>
<td></td>
</tr>
<tr>
<td>R5. How are participants’ approaching their FO course assessment practice?</td>
<td>7</td>
<td>Descriptive statistics</td>
</tr>
</tbody>
</table>

11.1 Participants’ Background Context

The following section provides the institutional, class\(^{30}\) and course contexts related to the second study interview participants. First, an overview of US higher education (HE) is provided. Second, the participants’ institutional contexts – US public two-year and public four-year HE institutions – are described. Third, differences in the nature of the F2F and FO teaching contexts are provided. Fourth, descriptions, as well as the wider context, are provided for the different courses used by participants in answering the interview questions.

11.1.1 Overview of US Higher Education

Upon successful completion of their secondary education, US students may end their formal education and, for example, go directly into the workforce or, more typically, pursue some form of trade or professional certification. This certification is typically offered by HE institutions that are either ‘two-year’ or ‘four-year’ (according to the typical length of study) and either public (i.e. government or tax-payer funded), private for-profit or private non-profit. As Table 25 shows, the teaching context for all the second study participants is the context where more than three-quarters of US HE student enrolments are concentrated and also where more than 80% of students take distance education courses. The following paragraphs present the institutional background – public two and four-year HE institutions – for the second study participants.

\(^{30}\) Commensurate with the language used by participants, the term ‘class’ or ‘classroom’ is used and not ‘lecture’ or, for example, ‘lecture hall’.
Table 25: Percentage distribution of US student enrolment in postsecondary education and in DE courses by institution type (2007-08)

<table>
<thead>
<tr>
<th>US Institution Type</th>
<th>Second Study Participants’ Context</th>
<th>Percent of Total US</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Student Enrolments</td>
<td>Students Enrolled in a Distance Education** Course</td>
</tr>
<tr>
<td>2 yr Public</td>
<td>P1*, P5 and P6</td>
<td>43.8</td>
<td>53.1</td>
</tr>
<tr>
<td>4 yr Public</td>
<td>P1*, P2, P3 and P4</td>
<td>32.1</td>
<td>27.7</td>
</tr>
<tr>
<td>4 yr Private not-for-profit</td>
<td></td>
<td>14.2</td>
<td>9.0</td>
</tr>
<tr>
<td>All for-profit (including &lt;2yr, 2yr and 4yr)</td>
<td>-</td>
<td>9.9</td>
<td>10.2</td>
</tr>
</tbody>
</table>

* P1 teaches at an HE institution that offers both two and four-year degrees. ** In 2007–08 a distance education course was defined as a course ‘taken for credit during the academic year that was not a correspondence course but was primarily delivered using live, interactive audio or videoconferencing, pre-recorded instructional videos, webcasts, CD-ROM or DVD, or computer-based systems delivered over the Internet’ (Radford, 2012, p.9). While it is unclear what proportion are FO courses, as defined by this thesis, other research suggests FO instruction is the most prevalent form of DE (e.g. Sumler, 2001).

11.1.2 US Public Higher Education Institutional Background Context

US public two-year HE institutions is the teaching context for P1 31, P5 and P6. Where these institutions are wholly focused on two-year HE instruction (i.e. for P5 and P6), they are commonly termed ‘community colleges’. After World War II, due to the high demand for and existing limitations of HE institutions, US community colleges became a powerfully dominant force in HE with ‘freshman enrollments at two-year community colleges and technical institutes [growing]…more than a fifteen fold…over three decades (1965-1980)’ (Thelin, 2004, p. 322).

These colleges generally became known to incorporate the following five components: ‘academic transfer preparation, vocational-technical education, continuing education, remedial education and community service’ (Cohen and Brawer, 1982, p. 15). They were initially built around and for the surrounding community (though the advent of FO instruction appears to be broadening these boundaries) and provided the opportunity for students to pursue some form of HE certification without the need, at least initially, to leave their home residence. Apart from meeting the increased demand for access to HE, these colleges were also known for their open access policy which welcomed all students to apply. In so doing, they welcomed many students who were academically underprepared and thus became increasingly important in meeting the progressively higher demand for remedial education – which universities and other colleges

31 P1 teaches at an HE institution that offers both two and four-year degrees. Henceforth, P1 will be considered as teaching in both the two and four-year HE contexts.
were either unprepared or unwilling to provide (Cohen and Brawer, 1982). Thus one of the major roles played by community colleges was the provision of remedial education that prepared students for transfer to four-year institutions (often termed the ‘2+2’ transfer program).

Community colleges offer students two significant educational tracks leading to an ‘associate degree’. They may enter a community college and enroll in a two or three year program leading to a trade certification such as computer network technician or nursing. Students may also enroll in a two-year transfer program that, with careful selection of courses, will be the equivalent of their first two-years of a bachelor’s degree in a public or private university or college (i.e. ‘2+2’ transfer program). In either case, students typically take (and instructors typically teach) five courses per semester that are typically32 worth three credits each. For students, these credits typically count towards the approximately 60 credits required to earn their degree. Most of these courses will be in their area of specialization (i.e. ‘major’), a small number may be ‘general elective’ courses that may be taken in other areas of specialization and others, such as mathematics courses, will be required.

To complete a two-year associate degree students generally need to complete approximately 20 courses, of which at least one or two are ‘required mathematics courses’. These courses may be particularly specified (e.g. introductory statistics is required) or broadly specified (e.g. using course numbers such as ‘MATH 120 and above’). A mathematics course(s) may also be taken as one of a small number (e.g. two to four) of ‘general elective’ courses that are typically required for degree completion. In a typical mathematics course sequence a student may take a ‘liberal arts’ mathematics course (covering, at an introductory level, a broad range of topics such as logic, set theory, linear equations and statistics) followed by an introductory statistics course. For example, a student studying criminal justice with eventual plans to transfer to a four-year HE institution may be required to take an introductory statistics course in addition to one more ‘college level’ mathematics course. They may also optionally take one or more mathematics courses as part of their general electives. As a result, given the typical diversity of programs offered at community colleges, a single F2F or FO mathematics course may have students from a variety of backgrounds and programs.

32 Courses range from counting one to four (or more) credits towards an overall degree credit requirement. For example, some courses, such as science courses that have a required laboratory, may count for four credits. Others, such as courses that instruct students on how to use Microsoft Office software, may be worth only one credit.
US public four-year HE institutions, the teaching context for P1, P2, P3 and P4, are commonly termed ‘state universities’. Like community colleges, they are tax-payer funded through the state government and also came into being in response to a demand for HE. However, unlike community colleges, their inception and growth started much earlier where the demand was for more practical rather than intellectual training that, in the US context, is associated with a ‘liberal arts’ education. This demand led to the US government’s 1862 and 1890 Morrill Acts which provided federal land to develop what were then termed ‘land grant’ colleges or universities (Lucas, 1994).

US public four-year HE institutions typically offer students a variety of undergraduate programs leading to a bachelor degree. As with associate degrees, the typical course is worth three credits and apart from courses in their area of specialization, students are typically required to complete some general elective and required mathematics courses, with the typical mathematics course sequence mirroring what may be expected at the community college level. A bachelor’s degree typically requires completion of 40 courses or about 120 credits, however, if a student is in a transfer program from a community college they may only be required to complete an additional 20 courses or about 60 credits – where all required mathematics course(s) may be completed at the community college level. For example, a community college criminal justice graduate, who completed a required introductory statistics course, is transferring to a four-year university program. At the four year level they will typically be required to take an additional 20 courses of which one may be a required mathematics course such as ‘advanced statistics’. Again, similar to community college instruction, a single F2F or FO class may also have students from a variety of backgrounds and programs.

Finally, there are at least two ways public two and four-year HE institutions may be differentiated. First, unlike instructors from public four-year HE institutions community college instructors are not required to engage in any research but are required to spend more time teaching. However, four of the six interview participants (P1, P2, P4 and P5) describe their ‘time in academia’ as ‘mostly teaching’, with the remaining two describing it as ‘about the same amount of teaching and research’ (P3 and P6). That is, P6 appears to be outside the norm where community college instructors are generally not expected to be involved in research. Second, because of open enrollment policies, the community college classroom is expected to be more diverse with a broader range of backgrounds, ages and abilities. For example, while many community college students may be coming directly from high school to study full-time, many may be studying part-time while working full-time and with families to care for (Mesa, Sitomer,
Ström & Yannota, 2012). However, one caveat is that some of these demographic differences may be less pronounced for FO students (Halsne & Gatta, 2002; Distance Education and Training Council, 2007), who are broadly considered to be ‘off-campus learners with a wide range of ages, work experience, and family circumstances’ (Mayadas, Bourne, Bacsich, 2009, p.86).

11.1.3 US Public Higher Education Mathematics Class Background Context

Apart from the institutional context, there are some differences in class contexts which the following section outlines. First, classes in two and four-year HE institutions are described. Second, a brief description of some fundamental ways the F2F and FO course contexts differ is offered.

Class size associated with both types of institution has been found to vary, for example, according to the level of mathematics (i.e. introductory vs. advanced) and type of mathematics (Kirkman, Lutzer et al., 2007). For example, the class section size of undergraduate mathematics F2F courses (up to Calculus) taught at PhD-granting four-year institutions (the context for P2, P3 and P4) ranges from 40 to 48 students. In contrast, the average class section size for these F2F courses taught at two-year institutions is 23 students (the context for P1, P5 and P6). These section sizes tend to be slightly larger for statistics courses (the course context for P1, P2 and P5; Kirkman, Lutzer et al., 2007). However, class section sizes for FO mathematics courses are somewhat more difficult to ascertain ‘because distance-learning sections are not bound by room-size limits and tend to vary dramatically in enrolment depending on local administrative practice’ (Kirkman, Lutzer et al., 2007, p. 149). That is, for example, participants may teach multiple sections together as a single class (e.g. a single FO mathematics course taught at a two-year institution may contain two sections or, on average, 46 students).

Finally, apart from differences in F2F vs. FO assessment practices, which were discussed in the literature review, some differences in US public two- vs. four-year HE mathematics classes may be expected. In particular, with smaller community college class sizes, the number and variety of assessment instruments may be greater and the feedback more individualized than what may be expected in larger four-year HE mathematics classes. Additionally, though expected in FO course contexts, the use of CAA systems may be even greater in four-year HE mathematics classes where the efficiencies they offer appear suited for assessing larger classes.
11.1.4 US Public Higher Education Mathematics Course Background Context

There are six different mathematics course contexts represented in the second study. ‘Introductory statistics’ is the course context for half of the participants and, for the remaining participants, the course contexts are varied (see Table 26). All courses are assumed to contribute three credits towards the overall degree credit requirements with the exception of ‘Mathematica for Calculus’ which P3 states is a one credit course. The following paragraphs describe, with respect to the US HE context, the typical content of these courses as well as the wider course sequence within which they are typically contained.

‘Introductory statistics’ is the F2F and FO course context for P1, P2 and P5. The typical US HE statistics sequence consists of an ‘introductory’ or ‘elementary’ statistics course which may be followed by a course in advanced statistics or probability. The typical introductory statistics course covers data collection and sampling, descriptive statistics (including graphical displays, measures of central tendency, variability and position), a brief introduction to probability and probability distributions, normal distribution and an introduction to inferential statistics.

‘Calculus I’, the F2F course context for P3, is a course taken by non-mathematics major students. According to P3, the majority of his/her students take this course because they are required to. The US calculus sequence is typically preceded by a course in pre-calculus which, upon successful completion, may be followed by taking, in order, courses in Calculus I, II, III and possibly IV. The typical calculus I course covers limits, differentiation and integration of one variable functions.

Mathematica in calculus, the FO course context for P3, is a course taught to students that are mathematics majors. The course teaches students how to use Mathematica software for calculus-level mathematics. The course begins by covering the basics of Mathematica software and includes use of Mathematica in defining functions, elementary equation solving, elementary graphing, limits, differentiation, logic and sets and integration. This course is one of just a few options students have in fulfilling a mathematics degree requirement for working with mathematics symbols in computer-mediated environments.

‘College algebra with applications’, the F2F and FO course context for P4, focuses on real-world applications of linear, polynomial, exponential and logarithmic algebraic models but is not intended for students planning on taking calculus level mathematics. To take this course,
students need a background in basic algebra (e.g. systems of linear equations, exponents, polynomials, solving quadratics by factoring, radicals and rational exponents).

‘Mathematics for Primary Teachers’, the F2F and FO course context for P6, covers, with particular relevance to primary mathematics instruction, real numbers, arithmetic operations, other number systems, set theory, algebra and problem solving. It may be followed by a second similar course covering more mathematics topics (e.g. geometry, probability and statistics) and may be preceded by a ‘liberal arts mathematics’ course covering, at an introductory level, a variety of mathematics topics such as set theory, logic and basic statistics.

Finally, within these courses, assessment schemes may vary considerably. In the US HE context ‘academic freedom’ means that instructors generally have significant control over how they teach their courses. This includes which assessment instruments they use, how they weight them and whether or not they use any invigilation. For example, one Calculus instructor may have several homework assignments worth 10%, several quizzes worth 10%, five tests worth 10% each and an invigilated final exam worth 30%. Another Calculus instructor may have five tests worth 10% each and an invigilated final exam worth 50%. Though, it is not uncommon for departments or institutions to specify that, for example, a final exam/project must be part of their assessment scheme and it must be worth, for example, a minimum of 30%.

11.2 Characteristics of Participants

All six participants (P1, P2...) teach mathematics in US public HE institutions. Their specific demographic and course background is detailed in Table 26. Table 27 details some participant characteristics found as part of the first study. The latter information is limited in that the first study data was collected about one year prior to the second study and the participants’ course context for the first study is not necessarily the same as their course context for the second study.
Table 26: Demographic make-up of interview candidates

<table>
<thead>
<tr>
<th>Participant</th>
<th>Institution Type (yr)</th>
<th>US Location</th>
<th>Enrolment</th>
<th>Experience* (yrs)</th>
<th>Survey Context*</th>
<th>Course Context for Interview**</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2 and 4</td>
<td>Northeast</td>
<td>3,500+</td>
<td>16+</td>
<td>Introductory Statistics</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6-10</td>
<td>FO</td>
<td>Same</td>
</tr>
<tr>
<td>P2</td>
<td>4</td>
<td>Southeast</td>
<td>10,000+</td>
<td>16+</td>
<td>Introductory Statistics</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6-10</td>
<td>FO</td>
<td>Same</td>
</tr>
<tr>
<td>P3</td>
<td>4</td>
<td>Northeast</td>
<td>22,000+</td>
<td>16+</td>
<td>Mathematica in Calculus</td>
<td>Calculus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-5</td>
<td>FO</td>
<td>FO</td>
</tr>
<tr>
<td>P4</td>
<td>4</td>
<td>Northeast</td>
<td>27,000+</td>
<td>16+</td>
<td>College Algebra with Applications</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-5</td>
<td>FO</td>
<td>Same</td>
</tr>
<tr>
<td>P5</td>
<td>2</td>
<td>Midwest</td>
<td>30,000+</td>
<td>16+</td>
<td>Introductory Statistics</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-5</td>
<td>FO</td>
<td>Same</td>
</tr>
<tr>
<td>P6</td>
<td>2</td>
<td>Southwest</td>
<td>60,000+</td>
<td>6-10</td>
<td>Introductory Algebra</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-5</td>
<td>FO</td>
<td>FO</td>
</tr>
</tbody>
</table>

*Based on the Study I survey data **Though reflecting the actual content, course names have been changed to preserve anonymity.

Table 27: Participant approach measures and PD information from the first study (increasing order of S&B measure)

<table>
<thead>
<tr>
<th>Participant</th>
<th>S&amp;B</th>
<th>CCSF</th>
<th>ITTF</th>
<th>FO related PD (Days)</th>
<th>Maths Focus in PD?</th>
<th>PD Workshops</th>
<th>PD Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>External</td>
<td>Internal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Travel Funds</td>
<td>Internal Grants</td>
</tr>
<tr>
<td>P1</td>
<td>14</td>
<td>20</td>
<td>30</td>
<td>3</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>P3</td>
<td>14</td>
<td>25</td>
<td>27</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>P5</td>
<td>22</td>
<td>29</td>
<td>24</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>P4</td>
<td>24.5</td>
<td>29</td>
<td>25</td>
<td>3.75</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P2</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>20</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P6</td>
<td>27</td>
<td>29</td>
<td>20</td>
<td>0</td>
<td>NA</td>
<td>No</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: 1. Sabbatical for FO training was not available for any of these participants. 2. ‘NA’ means ‘not available’

As shown, the final participant selection is varied. That is, for example, participants teach in a broad range of institutions (e.g. 2 to 4 yr from varied parts of the US), approach measures vary (e.g. S&B ranges from 14 to 27 out of a survey participant range of 12 to 31.5 and a possible range of 7.5 to 31) and most professional development backgrounds differ. Participants appear similar in that, for example, almost all have more than 16 years of F2F tertiary teaching experience, half use their statistics course for the focus interview context and almost all received financial incentives to develop their FO courses. Perhaps reflecting the teaching focus in community colleges, P5 and P6 have one of the highest CCSF and lowest ITTF measures, though only P6 has one of the highest S&B measures.

In all, three participants taught in four-year US public HE institutions, two taught in US community colleges and one taught in a US public HE institution that offered both two- and four-year degrees. For four of these participants, the course context selected for the interview
was the same as the course context used in the first study. One participant used a different course, but one they taught in both the F2F and FO contexts. The final participant taught only one FO course which they did not teach in the F2F context. In this case, a comparative F2F course was selected though the two courses were considered quite different in that, for example, this participant’s FO course focused on teaching mathematics majors how to use Mathematica while their F2F course focused on teaching non-mathematics majors Calculus.

11.3 Terminology Used in the Analysis

Several participants make reference to ‘MyMathLab’ (MML) and ‘CourseCompass’ (specifically referred to by P1, P2 and P4 and indirectly by P6). These are software products produced and sold commercially by Pearson Publishing. MML is described by Pearson as ‘a powerful online homework, revision and assessment tool designed to help students practise and improve their understanding of mathematics and to give their instructors feedback on their performance’ (http://global.mymathlabglobal.com). It is built on the CourseCompass course management system. Individual MML-based assessment questions may offer resources such as ‘Help Me Solve This’ or ‘View an Example’. Students may optionally use these resources to help them answer specific questions or solve specific problems being posed. MML may also be referred to as ‘courseware’, which is an abbreviation for ‘course software’.

Three additional terms appear frequently. First, throughout the interviews participants often use terms such as ‘online class’, ‘online course’ or ‘online setting’ when referring to their FO course. Second, the term ‘proctoring’ is used by participants. Its UK equivalent is ‘invigilation’. Third, the term ‘assessment feedback’ will be shortened to just ‘feedback’ for the remainder of the second study results. In doing this, the term is not to be confused with feedback given in the context of instruction (e.g. where an instructor is working on an example with the class). While assessment and instructional feedback may serve similar or over-lapping purposes, the interest in this study remains on feedback associated with weighted components that are part of a course assessment scheme.

Finally, the interviewer and six participants are referred to, respectively, as I, P1, P2, P3, P4, P5 and P6. In addition, some quotes appear multiple times as a natural consequence that some responses speak to more than one emergent theme.
11.4 Use of Discussion

Here discussion refers to dialogue occurring as part of a course, either between students or between the instructor and a student(s), which participants primarily intend to use as a means to help further student learning of mathematics. As many of the following quotes reflect, participants refer to at least four kinds of discussion: about the course itself (what P2 terms: ‘logistics questions’; e.g. course assignment due dates, clarifying questions), the mathematics being covered, assigned mathematics problems, and challenging open-ended but not necessarily assigned problems. One important distinction needs to be made between the first and second study. Whereas the first study focuses on the use of discussion as a weighted part of the overall assessment scheme, the second study expands the coverage to all discussion (i.e. ‘weighted’ and ‘non-weighted’)

Study I Background:

Findings indicate that fewer than half (39%) of participants use ‘discussion’ as a weighted assessment instrument (mean weighting of 10%). In addition, there appears to be no significant relationship between the use of discussion as a weighted assessment instrument and any of the approach measures.

Study II - R1: How and why is discussion/interaction used?

As the US HE background reflects, the second study participants’ mathematics classes are typically smaller than what may be found in some other countries where mathematics is taught in lecture halls to much larger groups of students. Therefore, these participants may be accustomed to having the opportunity for regular class discussion. However, with regards to its use as a weighted assessment instrument, only one of the second study participants reports using discussion in their F2F courses as compared to three in their FO courses (see Table 28). These numbers are consistent with the expectation that the F2F class setting provides a natural opportunity for discussion. They are also consistent with the first study where 39% of participants were found to be using discussion as a weighted assessment component in their FO courses.
Table 28: Summary of interview participants’ use of discussion as a weighted assessment instrument

<table>
<thead>
<tr>
<th>Participant</th>
<th>Weighting (%)</th>
<th>F2F Course</th>
<th>FO Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>P3*</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P4</td>
<td></td>
<td>-</td>
<td>4.5</td>
</tr>
<tr>
<td>P5</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P6</td>
<td>5.6</td>
<td>5.5</td>
<td></td>
</tr>
</tbody>
</table>

* F2F and FO courses were different for P3

This analysis is structured in three parts. First, participant-identified purposes for the use of discussion are detailed. Second, participant-identified differences are outlined and these are linked to reduced pedagogical capabilities in FO courses. Third, these differences are then linked to the overall quality of learning derived from the use of discussion in FO courses. While acknowledging some potential, the balance of evidence indicates a significant qualitative disparity in favour of the F2F context.

11.4.1 Purpose of Discussion

All participants appear to value the use of discussion in both course contexts. Three broad purposes emerge from the interviews.

First, course discussion is viewed as an opportunity for the instructor to monitor the state of student thinking and, if necessary, provide feedback that will help students stay ‘on track’ (P6):

\[ P2: \text{i monitor but i typically do not interfere unless i see something that is factually incorrect...} \]

\[ P4: \text{...trying to gain a little insight into their thinking.} \]

\[ P6: \text{i guess the goal is for me to say 'i recognize you're lost, now let's find a way for us to dialogue more to get you back on track'.} \]

Second, some participants identify the use of discussion as a means of building and maintaining community. They link discussion to collaborative learning, the ‘social construction’ of mathematical understanding and peer instruction:

\[ P2: \text{the first thing that i want to do is i want to create a community of learners with students in the course...you know, truthfully, this [i.e. discussion] is a great way for students to answer each} \]
other’s questions and me not having to spend the time on dealing with issues that students can really resolve before they need me to interfere.

P4: I believe very strongly in the social construction of knowledge, I like students to talk together. I believe that helps them learn better, so I was trying to create an environment for a little bit of collaboration.

Third, discussion is also identified for its potential as an aid to deepen understanding. In particular, it provides students with an opportunity to articulate their thoughts, either through writing or verbalizing:

P4: ... discussion boards have great potential in terms of getting at deep learning.

P6: ... in all of my classes, I ask these sort of reflective, broader kind of questions like ‘What is the math that we’re talking about? And, and how can you internalize that in a way that isn’t doing a bunch of procedures?’

In summary, discussion is viewed by these participants as a means of monitoring students and building and maintaining interactions. In addition, particularly relevant to the present research focus, it is recognized for its potential to bring about deeper understanding of mathematics.

11.4.2 Differences in the Use of Discussion

Participants identify several differences in how discussion is used. The first two differences present what may be considered initial state conditions for discussion in FO courses. The following four differences present what may be considered the resultant experience of FO course discussion, each reflecting reduced pedagogical capabilities in FO courses. Together, findings suggest that participants are struggling to use discussion in their FO courses. In particular, as will be discussed later, responses by both US community college interview participants (P5 and P6) appear consistent with an emphasis on teaching in community college instruction.

Differences in the Nature of Communication

Participants characterize F2F discussion as natural and FO discussion as unnatural. What appears fundamental to this difference is the nature of communication that contrasts the limitations of the largely unimodal (text-based) and asynchronous nature of FO communications to the potential of the more multi-modal (written, verbal, body language) synchronous nature of
F2F communication. P6, for example, contrasts F2F discussion that ‘naturally happens’ with FO discussion that does not ‘magically erupt’:

P6: ...So in my F2F classes...I go from group to group to group and I’m sort of able to challenge them... Whereas, in an online course, a) there isn’t dialogue just magically erupting because the students don’t necessarily see each other and b) I don’t have the opportunity to go from group to group to group and sort of connect those pieces so I use the discussion as a way to connect those pieces... [Later adding] Well, I feel that in an online class, there, there needs to be a dialogue that’s sort of happening. In a traditional classroom, that dialogue naturally happens in the classroom... in my online course, I always feel this desire... [for my] students to talk more – it’s like we’re missing, I’m missing that chance to sit down with them and just dialogue... sometimes I give feedback, but I don’t... feel like it generates the same sort of discussion... it’s like I build in some extra assessments to make sure that they really are getting what they’re getting cause I don’t have that intuitive assessment that happens when I’m sitting next to them.

P2: ...you know in a F2F class, we’re all so used to communicating with each other and if a student is maybe disruptive I can step in and handle it right then and there in real time. I spent some time on the front end in our [FO] course... I have an orientation video - a very brief one that I like to give students...[to explain expectations for engaging in FO discussion]... It’s a natural expectation [in the F2F course] and students have more experience in a classroom and they have a little bit more innate realization about what is and isn’t acceptable and if they start heading down that path, I can cut them off very quickly before it becomes damaging to the class.

In summary, these participants’ responses suggest there is something fundamentally missing with discussion as it is currently experienced in FO courses.

**Differences in How Discussion is Formalized**

Perhaps due to differences in the nature of communication, discussion in FO courses appears to be more formal (i.e. attached to a specific task or topic). This is compared to the greater potential, in F2F courses, for more informal discussion that is considered a benefit to student learning:

P2: I have class participation... that’s simply my judgement in the F2F... And then online I look at the discussion boards. I have some very specific things I ask them to do on the discussion boards... They are required to post three times [per chapter] and they have about four different kinds of things that they can do: they can post what is the most helpful, what’s the least helpful thing, they can post a question and they can answer a question.
P3: [Comparing text-based FO discussion to verbal F2F discussion] I don’t think I can write an essay, I am a pretty good writer, but I don’t think I can write an essay that’s as wide-ranging as a discussion I can have with a math major... when I write an essay, I can’t go off on a tangent; whereas when I’m talking with a math major, I can go off on several tangents, and some of them are very fruitful.

P5: ... I can give them informal feedback in a F2F class... we have more informal discussions. Whereas in the online class that’s not really possible...

P6: In my online course, in order for them to get full credit [i.e. a grade that counts towards the discussion weighting], they have to respond to each other because they need that dialogue, but in the F2F class we get that dialogue in the classroom...you know my students will tell you that I get off topic all the time... I allow the conversation to get off topic in the classroom sometimes because I think it’s valuable educationally for us to have those little side notes about how does this connect to something else you’re thinking... And it’s like we miss those in an online course.

In summary, FO discussion appears to be more constrained than F2F discussion. And as two participants (i.e. P3 and P6) state, FO students miss out on the value of more open-ended discussion. The remaining differences present four inter-related concepts, each pointing to reduced pedagogical capabilities in FO courses.

**Difference in the Amount of Communication Resources**

With the loss of direct F2F contact, two participants identify a reduction in multi-modal communication – specifically, the loss of verbal and visual communication that makes it more difficult to monitor and communicate with FO students. Instructors in the F2F context, for example, can monitor student understanding simply by ‘reading faces’. However, in a FO course, the onus – that involves both recognizing and then communicating in writing – is placed on the student to communicate any difficulties they may be experiencing:

P2: Of course, in an on-ground course, you can quickly verbalize your expectations and say things very quickly that a F2F class understands where it’s much more difficult do that in an online course

P6: I really feel like there’s a lot more to communication than just text and just words... even very mathematical text misses the, all the other parts of communication, it misses the gesture, it misses the graphic...[versus in FO course] in the WIMBA [i.e. web-based software that enables synchronous interactivity] classroom, they have a little whiteboard that they can draw on with their mouse and they can also show each other their screens, so they can put together a power
point slide and say, ‘No, no, I mean this!’, you know. But it’s not as effective, there’s something really valuable about F2F communication that is lacking... I’ve yet to find a way to construct an activity that would allow for that kind of dialogue or exchange to take place that would be as effective as five minutes in the classroom.

In summary, two participants identify the loss of direct F2F contact in FO courses with a loss of resources for communicating. The result is that what may be communicated in a relatively short time in a F2F course discussion, may take considerable time in a FO course discussion.

**Differences in the Use of Instructional Timing**

Participant-identified differences contrast both the potential and problems with the use of instructional timing in FO courses. Instructional timing is broadly defined as the timing associated with any aspect of instruction. That is, for example, when to start covering new material, when to ask a new question, when to give a hint to a previously posed question or when to end that question and begin explaining.

As previously discussed, the use of discussion was identified for its potential in deepening student learning. Compared to the F2F context, some participants identify a greater potential in the FO context where the asynchronous nature of discussion appears to remove time pressures and thus enable students to be less inhibited, more reflective, and to articulate and write down their thoughts about mathematics:

**P6:** I’ve had some students who have been in my F2F class and then they’ve gone to one of my online sections like for the next semester and it’s really interesting because I would have pegged them to be kind of a quiet student in class and then they write these beautiful, big flowing page-long discussion posts and I’m like, ‘aha!’ This is a student who just needed some time to think or they just needed some space to be able to share their ideas and that wasn’t available to them in the F2F context.

**P4:** ... by writing their thoughts down it forces them to be more thoughtful; so I think the discussion boards have great potential in terms of getting at deep learning...

Extra time provision in FO discussion appears to be particularly noted with respect to how students are better able to reflect on peer comments and provide peer instruction:

**P6:** I think there’s something nice about forcing students - in this asynchronous discussion, they really have to listen to what the other person wrote, they have to read it to be able to respond. And so there’s this... sense... that students are actually attempting to understand each other
sometimes and not jumping over ‘No, no, let me show you my idea’… So there’s a beauty in that kind of discussion.

P2: You know, truthfully, this is a great way for students to answer each other’s questions and me not having to spend the time on dealing with issues that students can really resolve before they need me to interfere… Peer instruction is very, very helpful…

However, two problems are identified, one of which appears significant. First, contrasting with students benefitting from having more time to, for example, respond to questions in FO courses, some participants see themselves as having less time (i.e. needing to respond ‘quickly’) to respond to their students:

P2: The [FO] discussion board, the [FO] homework discussion board can help fill the gap a little bit, but if there’s a real issue or concept of something that needs some attention, I just want the student to have attention very quickly.

P5: …in the online class… I want to make sure the students get the formal feedback as quickly as possible and learn from their error while things are fresh… if I delay in returning assessments in the online class, I think the students really do lose a lot from that.

Second, in what appears to be a significant issue, the potential precision and efficacy of instructional timing in the F2F context is contrasted with an unknown precision and efficacy in the FO context. The following statements highlight the potential role of discussion that is ‘real time’ (P5) where even timed pauses of silence are used as an important tool for helping further student understanding. Both are seen to be part of F2F but not FO courses:

P3: There’s a tremendous difference; in the sense that when I’m teaching [F2F] calculus, I pause – I ask a question – and then I pause, and I give them a couple minutes to answer it. And then, then I say ‘Has everybody finished?’ And then I sometimes ask students to come to the board and explain their answer, which is not something I can do in an online course.

P5: The fully online class… I don’t have the opportunities to scaffold them as easily in real time.

In summary, participants view the nature of instructional timing in FO courses as potentially beneficial to student learning. However, these benefits are offset by identified problems with executing effective instructional timing which, when considered along with previously discussed differences in the nature of communication, suggest it is more difficult in FO courses to both judge and execute appropriate instructional timing.
Differences in the Use of Interpersonal Dialogue

Differences in the use of interpersonal dialogue present another example of emergent differences in pedagogical capability. Interpersonal dialogue is described by Caspi and Gorsky (2006) as, structurally, ‘a message loop between Instructor–Student–Instructor or Student–Instructor–Student or Student A–Student B–Student A’ (p.737). Relatedly, in their FO courses, participants identify the loss of an iterative or cyclical instructional dialogue (e.g. student-instructor-student...) that occurs in the moment and is considered necessary for effective instruction. Such a loss is suggested by P3’s previous statement. The following statements provide further clarity:

**P5:** Now I do that sort of thing [i.e. engage in a conversation about mathematics] in the online class, but it’s usually asynchronously, so I’ll ask a question, the student will respond an hour later and then we’ll go from there, so I don’t think that’s nearly as effective as it is an a F2F environment. Mathematics is a language, mathematics, no question, is a conversation... So whenever you’re doing feedback, the feedback should go both ways... I talk to a student, the student talks to me...and from that conversation... I think students can develop understanding.

**P6:** But it’s lacking... that synchronous give and take that happens in the classroom where in the classroom, when the student puts that up and they get to the third line and they go ‘Oh wait - I don’t know how do that’ and I say well ‘Oh, go ahead and finish putting that up and let’s talk a little bit about that mistake and why you thought that was a mistake’. And so I get to kind of sit over their shoulder and give some dialogue that I don’t get to do in an asynchronous discussion... I always feel this desire in my online course to like, I want the students to talk more – it’s like we’re missing, I’m missing that chance to sit down with them and just dialogue... that dialogue doesn’t always happen when they’re just sending a written assignment and sending it in and sending it in...

As the only community college participants these responses from P5 and P6 may be expected. That is, it can be argued that there is more of a culture of discussion (and interaction) in US community college classrooms given class sizes, on average, are smaller than those in US four year HE institutions and the emphasis is on teaching, not research. In this regard, it may be expected that these participants are more sensitive to any restrictions on the use of discussion presented by the FO environment.

Along with the differences in the use of instructional timing and interpersonal dialogue the situation may be better understood by the following summary description: In F2F courses, students and the instructor are expected to be physically in the same place at the same time.
Communication may be both verbal and written and there can be a real-time cyclical form of dialogue that attempts to combine the effective use of instructional timing and interpersonal dialogue. However, in FO courses, students and the instructor(s) are physically separated in space and time. Communication may only be text-based and the same ability to effectively combine the use of instructional timing and interpersonal dialogue appears to be missing. P6 sums up the situation:

\[ P6: I \text{'t} \text{ have the flexibility [in the FO context]... [while in the F2F context] I can assess their construction by sitting down with a group and listening to their conversation and talk back and forth with them...} \]

In summary, in their FO courses, some participants experience a felt loss of real-time ‘give and take’ (P6) that is a characteristic of their F2F discussion and instructional practice. These accounts suggest the current nature of interaction in FO courses is not providing the necessary conditions for effective mathematics instruction. In particular, the current FO course environment is not facilitating the tight nature of the ‘message loop’ (Caspi & Gorsky, 2006), or iterative feedback loop (Hounsell, McCune, Hounsell, & Litjens, 2008), that instructors need to help develop students’ understanding of mathematics. These issues will be further addressed in the coming sections on the use of feedback.

**Differences in the Use of Collaboration**

Bringing together many previous issues is the use of collaboration. While every participant, with the exception of P3, appears to use collaborative activities in their F2F courses, of these five all expressed difficulty using collaboration in their FO courses. Referring to their FO course experiences in attempting to use collaboration, the following quotes are illustrative:

\[ P1: \text{...interaction [in FO courses] seems to be very difficult for students.} \]

\[ P4: \text{It’s much more difficult for me to encourage collaboration online [in FO courses]; I’m not so sure I’ve succeeded very well. But, in F2F I just have them do group work in class and you know have them do it outside of class and it’s much easier.} \]

\[ P5: \text{Well, in my F2F class we do several collaborative exercises. I have students work on problems together, I encourage students to get up and move around the classroom to discuss strategies, to check solutions, to argue and debate about types of problem-solving strategies. So that’s a significant part about how I do my F2F classes and I have not yet been able to figure out} \]
how to do that in an online [FO] environment. I have not been successful trying to translate that into that environment...

P6: [Compared to F2F], in an online course, a) there isn’t dialogue just magically erupting because the students don’t necessarily see each other...

Even though some students may be accustomed to collaborating, particularly those studying on-campus but taking FO courses, some participant responses suggest the idea of collaboration outside of the ‘time structure of a F2F class’ is not feasible:

P1: [The reason FO discussion does not work is] [simply the difficulty of the logistics online, I have students from around the world... between the time difference and the fact that everybody’s schedules seem to be very different...

P4: I think maybe people that take an online class like to work at their own pace when it’s convenient for them as opposed to the time structure of a F2F class, I think that has something to do with it [i.e. problems with FO collaboration].

In particular, as the following quote states, some students do not want to collaborate in their FO courses:

P5: I have tried to encourage [FO] students to work together on projects... and it did not go very well, the students were extremely resistant. I had emails and phone calls saying things like this: ‘This is an online class, you shouldn’t expect me to work with my classmates’ and ‘We can’t find time to work together’ and ‘We do not want to get together in person’ - even though they were within 3 or 4 miles of each other ‘We want to do things electronically’.

In summary, while it appears that almost all participants recognize and practice collaboration in their F2F courses, they are struggling to do so in their FO courses.

Overall differences suggest the use of discussion in FO courses is unnatural and more formalized with several differences suggesting overall diminished pedagogical capabilities.

11.4.3 Quality of Learning and the Use of Discussion

Taken as a whole, participants’ accounts of the use of discussion suggest a toll on the quality of FO mathematics assessment and instruction. Directly related to the quality of learning, participants identify two additional issues. First, some responses describe how easy it is, in FO courses, to use someone else’s work and pass it off as your own:
P4: ... one of my colleagues stopped using discussion boards because they just copy what the other person wrote... just to get the points they just copy and paste and change a little bit. I told them they had to change, that they had to use a different example...

P5: ... then when students were trying to collaborate electronically, that was not going very well. I was hearing that one student was doing the collaborative exercises and submitting them for the others...

Second, despite any stated potential or questions regarding a lack of pedagogic strategies for the FO context, most participants do not associate FO course discussion with quality learning. Instead, many believe that only F2F discussion is capable of effectively helping students learn at a deep level:

P3: I think for students, for math students to understand mathematics at a deep level, there has to be a lot of F2F.

P4: ...you know the quality work [in discussions] wasn’t there - and so I’m not saying that you can’t do it, but it would probably take probably a little bit more work... you think with all the communication tools available on the internet, you know this communication would be cool but its hard time to get them to interact together online.

P5: The fully online class I have difficulty preparing those students for open-ended questions [Then going on to attribute this to both the lack of collaborative activities and the lack direct F2F contact needed to effectively ‘scaffold’ students’ thinking].

In balance, most participants seem to express the need – what for some almost seems a longing – for live F2F interactions with their students, and struggle with the largely text-based and asynchronous communication offered in FO courses. Reflective of these struggles, many are opting out of using discussion as a component of their FO instruction (those noted here, by taking stock of student opinion) while others appear to be attempting to mediate for these challenges by encouraging students to use, where possible, live F2F office hours33:

P1: I polled the [FO] students in about a ‘you-must-respond-to-this-email’ and the unanimous response was that it [discussion] would just be a waste of time, which I thought was very interesting...We do do some group work in [F2F] class, small groups, the typical 3-5 sort of

33 Only one quote regarding the use of office hours is offered here. This issue will be discussed in greater detail in the section on the use of feedback.
thing. We do a lot of class discussion... [Then bluntly, in response to how, if it all, interaction is encouraged in FO courses] I don’t.

P5: I was asked [by students] to give up on the collaborative exercises and let them do them individually – and I agreed.

P6: And often, my solution to that [i.e. problems with FO discussion] is not a very good one because my solution is, ‘Please come meet me in my office hours, let’s have F2F dialogue’...

In summary, the use of discussion in FO courses appears to be a prominent issue. Where all participants describe the use of discussion as a natural component of their F2F instruction and assessment practice, in their FO courses, most participants are struggling to adapt to the FO environment while others are opting out of using discussion altogether. The emergent reasons for this disparity contrast the potential of using synchronous F2F discussion for developing students’ understanding with a struggling and seemingly disadvantaged asynchronous FO discussion.

11.5 Use of Quizzes

As discussed in the literature review, quizzes are defined as short oral or written tests (www.merriam-webster.com). As will be shown, participants characterize them as a well-known assessment instrument largely completed in a written or typed CAA input format, though used differently in different contexts.

Study I Background: Findings indicate that just over half (53%) of participants use quizzes in their FO courses (mean assigned weighting of 20%). For those who use quizzes as a weighted assessment instrument, the weighting allocated to quizzes was found to be correlated with the degree to which an information transmission/teacher-focused approach to teaching is used.

Study II - R2: How and why are quizzes being used?

As previously discussed, with regards to the US HE context where mathematics classes are typically smaller than what may be found in some other countries, second study participants’ appear accustomed to using quizzes as part of their class instruction. However, in the second study, as Table 29 summarizes, despite all participants using quizzes as a weighted assessment component in their F2F courses, only three use quizzes in their FO courses. The latter number is consistent with the first study where 53% of participants were found to be using quizzes as a weighted assessment component in their FO courses.

191
Before results of the analysis are presented, the use of CAA is discussed because its use emerges as an important characteristic of FO mathematics assessment practice. Following this, the analysis of participants’ accounts of their use of quizzes is structured in four parts. First, participant-identified factors associated with the use of quizzes are detailed. Second, participant-identified purposes for the use of quizzes are detailed. Third, participant-identified differences are outlined and these are linked to reduced pedagogical capabilities in FO courses. Fourth, the balance of findings is discussed and the use of quizzes in the FO context is linked by participants with lower level learning. Finally, there is one important limitation concerning participant-identified views regarding the use of CAA. As previously discussed, because it appears the majority of participants use one single CAA system (i.e. MML), their views may reflect more on that system than the use of CAA in general.

Table 29: Summary of participants’ use of quizzes in each context

<table>
<thead>
<tr>
<th>Use of Quizzes</th>
<th>F2F Course</th>
<th>FO Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAA/Paper-based*</td>
<td>Weighting (%)</td>
</tr>
<tr>
<td>P1</td>
<td>Paper</td>
<td>14</td>
</tr>
<tr>
<td>P2</td>
<td>CAA</td>
<td>10</td>
</tr>
<tr>
<td>P3**</td>
<td>Paper</td>
<td>10</td>
</tr>
<tr>
<td>P4</td>
<td>Paper</td>
<td>10</td>
</tr>
<tr>
<td>P5</td>
<td>Paper</td>
<td>20</td>
</tr>
<tr>
<td>P6</td>
<td>Paper</td>
<td>5.6</td>
</tr>
</tbody>
</table>

* Here ‘CAA/Paper-based’ refers to whether the quizzes are generated and graded by computer or directly by the instructor.
** F2F and FO courses were different for P3

11.5.1 Evidence for FO Course Dependence on CAA

In both thesis studies the use of CAA emerges as an important characteristic of FO mathematics assessment practice. In particular, interview participants’ responses suggest that the use of CAA has a significant influence on the quality of learning. Moreover, because it is assumed that FO courses make use of CAA more than F2F courses, this is seen to imply that its influence is greater in FO than F2F courses. The assumption that FO courses are more dependent on CAA than F2F courses is based on various sources both within and outside the thesis studies:

First, evidence of a greater dependency is reflected in some interview responses:

P5: ... [In FO courses] many people just want to use the packages [CAA] that are put forth by the publishers and they, you know, sacrifice some of these higher level skills for ease in terms of grading and implementing a course. So it’s a real problem.

P4: Of course...in the online course the quizzes and exams are all on the online software, which is MyMathLab; and in the F2F, of course, it’s paper and pencil.
Second, each of the participants’ F2F and FO course assessment schemes were compared. Based on this comparison (see Table 30) it can be seen that the majority of participants either use the same amount or more CAA in their FO than their F2F courses. This is further consistent with quiz use, shown in Table 29, which indicates almost all interview participants used paper-based quizzes in their F2F courses and, of the three participants using quizzes in their FO courses, only one uses paper-based while the other two use CAA-based.

Table 30: Participant use of CAA in F2F vs. FO courses

<table>
<thead>
<tr>
<th>Participant</th>
<th>Assessment Weighting Associated with the Use of CAA*</th>
<th>Course Using CAA the Most (difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>FO (100)</td>
</tr>
<tr>
<td>P2</td>
<td>20</td>
<td>same</td>
</tr>
<tr>
<td>P4</td>
<td>8</td>
<td>FO (79.5)</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>same</td>
</tr>
<tr>
<td>F2</td>
<td>0</td>
<td>same</td>
</tr>
</tbody>
</table>

*P3 is not included because the courses, referred to in the interviews, were not the same.

Third, perhaps the most compelling evidence, the first study survey findings (n=66 courses) on the use of CAA in FO courses was compared to similar findings from the latest available US Conference Board of Mathematical Sciences (CBMS) survey findings for F2F courses (n=600 mathematics programs, Kirkman, Lutzer et al., 2007). To do this, the first study survey participants who indicated they used immediate and/or computer-generated feedback were considered to be using CAA and these numbers were compared to the CBMS findings. For two-year institutions, in line with the CBMS survey definition, any use of CAA was considered. For four-year institutions, in line with the CBMS survey definition, only the use of CAA for homework was considered. As Table 31 shows, the percentage of the first study survey participants using CAA for their FO course homework clearly exceeds the percentage from the CBMS 2005.

34 Part of the 2005 CBMS survey asked: “departments about the use of a new teaching tool in their first-year classes, namely the use of online homework and testing software that was offered by many textbook publishers (and others) in fall 2005. The two-year questionnaire described these online systems as using ‘commercial or locally produced online-response homework and testing systems’, and the questionnaires sent to four-year mathematics and statistics departments described them as ‘online homework generating and grading packages’.” There are two limitations to this data: First, it was gained by asking department heads, not actual instructors, to estimate usage. Second, the use of CAA is expected to have increased since 2005 (e.g. characterized by Kirkman as having ‘exploded’; personal communication, August 21st, 2012). Despite these limitations it is doubtful usage has ‘exploded’ to the extent of the differences shown in Table 29.
Table 31: Comparison of F2F and FO course usage of CAA for homework

<table>
<thead>
<tr>
<th>Courses</th>
<th>Percent of Participants Using CAA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study I for Homework</td>
</tr>
<tr>
<td>2yr</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>80% (28/35)</td>
</tr>
<tr>
<td>Elementary Statistics</td>
<td>70% (7/10)</td>
</tr>
<tr>
<td>Pre Calculus, Calculus I, II and III</td>
<td>72.7% (8/11)</td>
</tr>
<tr>
<td>4yr</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>38.7% (12/31)</td>
</tr>
<tr>
<td>Elementary Statistics</td>
<td>33.3% (2/6)</td>
</tr>
<tr>
<td>Pre Calculus, Calculus I, II and III</td>
<td>41% (7/17)</td>
</tr>
</tbody>
</table>

*Average across 28 categories of courses. **Not available.

11.5.2 Identified Factors Associated with the Use of Quizzes

Participants identify several factors related to the use of quizzes. These factors include the nature of the instructional context, the nature of mathematics and the availability of appropriate resources.

First, aspects of the F2F and FO course context are identified as reasons why quizzes are used in F2F but not the FO course context:

*P1: [I: Why are quizzes used in your F2F but not your FO course?] I use quizzes in the on-campus class F2F because I can’t afford to have seven tests in the semester [i.e. these quizzes take the place of three extra tests, which would require the use of extra class time, that are administered in his/her FO course].*

*P3: [I: Why are quizzes not used in your FO course?] ... there’s no context for quizzes in an online course. Because there’s no time limit... I don’t use any of the current techniques for time limiting them. [In the FO course] they’re given a problem set [i.e. as homework]. They get it two weeks before their answers are due, and I collect their answers two weeks later.*

Second, the complexity of mathematics being assessed is also identified as a factor:

*P3: [In further response to why quizzes are not used in his/her FO course] No, in my online course, it’s completely not a computer assessment, because of the nature of their course. They’re being asked to answer the problems, which are pretty solvable problems in mathematics. They’re being asked to verify complicated computations.*

*P6: ...there is no online homework system available for the text book I use and there shouldn’t because the questions aren’t really very procedural questions, they’re very thought-provoking questions; and so there wasn’t really a way for them to be automatically graded.*
Third, following on from the last quote, as P6 alludes, s/he was unable to make use of quizzes because the appropriate resources were not available (in this case, appropriate CAA-based quiz):

\[ P6: \text{No, I don’t have quizzes in the online course for teachers; I do have quizzes in my other online courses... \[I: \text{Reasons?] It’s going to sound really lazy: In the other courses, I’m able to use an online homework system... In the online teacher course... there is no online homework system available for the text book I use...} \]

Based at least in part on these factors, all six participants state that they use quizzes in their F2F courses while only half use quizzes in their FO courses (see Table 29). Again, these numbers are consistent with the first study where 53% of participants were found to use quizzes in their FO courses. In addition, the contrast in numbers using quizzes in F2F versus FO courses suggests these participants view the use of quizzes as less feasible in FO than in F2F courses. This issue will be further explored in the coming section on differences in the use of quizzes.

11.5.3 Purpose of Using Quizzes

Across both contexts, two major purposes for quizzes emerge from the interviews. Surprisingly little is said relating quizzes to student learning. Instead, the two recurrent themes regard monitoring and directing student activity.

First, monitoring is seen as a means of informing the instructor if the instruction needs to be altered:

\[ P2: \text{I use them [i.e. quizzes] to figure out what do I need to do next.} \]

\[ P4: \text{[Quizzes are]... also to help them know what they know and help me know [what] they know and can do.} \]

\[ P5: \text{[Quizzes results help me]...so that I can go back and re-teach something if large numbers of students are struggling... one way for me to monitor whether or not the students are doing what they’re supposed to be doing, and it gives some additional feedback to try to find out if there ’s problems where those problems are.} \]

\[ P6: \text{The quizzes are a chance for me to actually say ‘You either understood it or you didn’t’.} \]

Second, quizzes are viewed as a means of directing students through the course material in at least three ways: engagement, pacing and transitioning. With regards to engagement, quizzes are viewed as a means:
P1: ...of keeping students focused on their work...and help[ing] them stay with [the instructor].

P3: ...to make them do their homework.

P4: ...just get them doing more math.

A comment left by one participant (other than P1 to P6) with their response to the fifth S&B question on the Study I survey: ...encourage them [i.e. students] to regularly get involved in learning the material.

With regards to pacing, quizzes are viewed as a means:

P2: ...to pace the course so that we can pay attention as a class to what’s coming up and what we need to be prepared to do.

P3: ...to keep them on their toes.

Where quizzes are viewed as a means of helping students make transitions, three kinds of transitions are identified. P1 refers to the transition from secondary to tertiary education. P4 and P5 refer to the transition from concept to concept within a course. Finally, the most common transitional purpose, as exemplified in the last three of the following quotes, suggest that quizzes are intended to act as a transition from or to another assessment instrument:

P1: ...ease their transition from the idea in high school that you have a test every Friday, to the idea in college that you have many fewer major exams.

P4: [In group quizzes] I will do something a little harder or something I have just alluded to as a segue to the new materials.

P5: I use the quizzes as a transitional element; so if my next topic is going to require some prerequisite knowledge, I might throw that in a quiz, or I might use the quiz to do some problem-solving, ask them to solve a problem they haven’t seen before related to the coming lecture.

P3: The quizzes are really meant to be preparation for the exams.

P2: [Quizzes help us know] what we need to be prepared to do in terms of assessment.

P6: [Quizzes are] usually two or three questions directly from homework.

Combining many of these purposes, this response from P2 illustrates how data, from CAA-based quizzes, is being used to both help monitor and direct student learning:
P2: I provide feedback on their quizzes in an aggregate fashion; I look at the most missed topic and make something that is accessible, in terms of feedback, to all the course. I also do coursecompass homework, which is repeatable as many times as they want until the due-date. If I notice a student that is repeating the question multiple times, I will interfere and give a little bit of direction.

Finally, in one of the only mentions related to learning:

P4: [The purpose of quizzes is] to kind of solidify their knowledge about something - knowledge, understanding and/or skills.

In summary, advancing student learning does not figure prominently. Instead the overall emphasis appears to be on monitoring and directing student activity.

11.5.4 Differences in the Use of Quizzes

This section summarizes how participants view the use of quizzes in FO as compared to F2F courses. Though some participants report little difference in how quizzes are used in either context, overall, at least four differences are identified. First, participants identify differences in how quizzes are used and valued. Second, participants identify differences with respect to how quizzes are administered. Third, quizzes are monitored differently in each context. Fourth, one participant identified a difference in the way feedback is provided. Overall, despite some identified potential, participants describe the FO course context as a more challenging environment which, in effect, either prohibits or inhibits the effective use of quizzes.

First, three participants highlight the different ways quizzes are used and appear to be valued in one or the other course context. For example, P2 appears to be using quizzes in his/her FO courses to mediate for challenges in pacing students. P4 views his/her F2F paper-based quizzes as more capable of challenging and helping to develop students’ learning – a difference s/he relates to a dependency on CAA in his/her FO course. On the other hand, while P6 appears to be using quizzes for monitoring student understanding in his/her F2F courses, s/he uses FO discussion to monitor understanding in his/her FO course:

P2: Yes, I think it’s [i.e. use of quizzes] more important in the online setting. I think it’s very difficult to set a pace in an asynchronous course and so that’s just one of the things I use to help set the pace.
P4: In the F2F there’s probably a little broader purpose and sometimes, it’s to challenge to go beyond their level of thinking as to where they are... It’s very difficult on a computer [i.e. using CAA in their FO course] to gather... to test conceptual understanding or, ability to problem solve...

P6: [I: So what you hope to accomplish in your F2F courses with quizzes is verifying by sampling some of the homework questions to verify that they understood the homework?] Yes. [I: Whereas, in the fully online context, you don’t feel that the need is there because they actually submit the homework and you actually grade the whole homework?] I’d like to tell you that’s true but often times, I don’t grade their whole homework: I do usually grade one or two questions off their homework so in some sense, it’s similar. But I get this great homework discussion happening, so I actually know what parts of the homework they don’t understand ‘cause they just flat-out tell you in the discussion. So in some sense, I’m getting that verification about what they don’t and do know in the discussion section rather than having to grade their homework.

Second, consistent with the previous discussion on how smaller class sizes make the use of collaboration more feasible, two participants identify differences in the way quizzes are administered which contrast the lack of flexibility in FO courses with greater flexibility in F2F courses. That is, in F2F courses, participants have the choice to administer quizzes either to be completed individually or collaboratively in small groups, whereas, in FO courses, responses suggest this choice does not exist and the expectation is that they are completed individually:

P4: And in the F2F I often, but do not always, have them do group quizzes...when I give a group quiz...

P5: Yes, in the F2F class, quizzes are primarily collaborative, occasionally, I have the students work on them on their own, but more often than not, I have the students collaborate...

Third, three participants’ responses suggest that FO course quizzes are not invigilated. Responses allude to the expectation that students will cheat as well as the hope that they won’t:

P3: [In the context of discussing his/her use of CAA-based quizzes] I would like to only be concerned that they only do the work correctly. But unfortunately, that’s not possible. Since, if I gave them take home exams, they would all cheat...

P4: And in the online environment, the purpose is similar; and those are not proctored so I am aware that they cannot follow my guideline, which is please try do these without using any outside assistance.
P5: The quizzes are monitored in the F2F class, they’re proctored. They are not monitored in the online class.

Fourth, one participant appeared more likely, in his/her FO course, to provide feedback via a recording:

P2: In a F2F class, I may go back to class and say ‘Hey, it looks like everybody struggled with regression, so I’m going to give you a brief overview and I’m going to give you some resources.’ Whereas in an online class, I probably would do something similar or a screen recording or a brief little video recording as feedback to the quizzes.

These last two issues, respectively, will be covered in greater detail in the coming sections on invigilation and feedback.

In summary, identified differences relate to how quizzes are valued, administered and invigilated as well as how feedback is provided. These differences suggest that the FO course context is a more challenging environment for the effective use of quizzes and provide one reason why, though all participants use quizzes in their F2F courses, only half use them in their FO courses.

11.5.5 Quality of Learning and the Use of Quizzes

Participants’ responses suggest that quizzes are used more as a tool related to completing course tasks (monitoring and directing student activity) than learning course content. While little is said that directly relates quiz use to student learning, emergent differences suggest the FO course context is a more challenging environment to realize quality learning through the use of quizzes. Participants’ responses suggest at least two principal reasons: First, as previously discussed, the FO course context appears to lack some pedagogical capabilities, present in F2F courses, which have been associated with the quality of learning (e.g. use of collaboration and invigilation). Second, in what will now be discussed, the reliance on CAA in FO courses, in general, and for quizzes, in particular, is identified by participants with lower-level learning. However, these views are considered limited by the predominant use of MML.

The greater reliance on CAA systems in FO versus F2F courses is consistent with participants’ accounts (when used, five out of six F2F course quizzes are paper-based while two out of three FO course quizzes are CAA-based; see Table 29). Participants associate CAA-based quizzes
with repeat attempts and small incremental steps in learning, where that learning is facilitated by questions that are limited to lower-level understanding.

The following responses illustrate how CAA-based quizzes are associated with repeatedly attempting the same or similar questions. P3 casts this process in a positive light while P5 and P6 consider it as detrimental to student learning:

\[
P3: \text{[The computer] tells them right away whether they have done a problem wrong... [if they got it wrong] that cycle repeats until they can do the problem right...}
\]

\[
P5: \text{...in the courseware they can hit buttons that say ‘Help, help, help, help, help’...}
\]

\[
P6: \text{[Students] do the same procedure over and over and over again...}
\]

The association with learning in small increments is noted by two participants:

\[
P5: \text{[The content focus is on] small packages.}
\]

\[
P1: \text{[Quizzes focus on] little bits of topics.}
\]

Lastly, the following quotes refer to the kinds of questions addressed with CAA as ‘trivial’, doing ‘computations’ or ‘procedural’ and not ‘thought-provoking’:

\[
P3: \text{[I: What do you hope to accomplish by using quizzes?] ...We have an online homework system, but it can only ask the most trivial questions, it can only ask them really to do computations.}
\]

\[
P4: \text{It’s very difficult on a computer [i.e. using CAA]... to test conceptual understanding or, ability to problem solve...}
\]

\[
P5: \text{... they tend to ask very low-level skills... basic knowledge questions...}
\]

\[
P6: \text{[explaining why s/he is not using CAA] ...because the questions aren’t really very procedural questions, they’re very thought-provoking questions.}
\]

Moreover, expanding on previous responses and limited by the predominant use of MML, CAA systems appear to be providing the scaffolding but not the learning. That is, there is a sense that once the scaffolding provided by CAA systems is removed the constructed knowledge does not remain standing. Instead, it appears students develop a dependency on CAA-provided scaffolding without necessarily developing the understanding. This suggests further concerns about lower-level learning:
P4: ...with MyMathLab or some of the, most of the online products; you know it’s too easy to get assistance, it’s ‘Help me solve this’ or whatever. I want them to try and see if they can do it on their own.

P5: [Students complete repetitive CAA questions] without necessarily internalizing [the mathematics]...

In summary, FO course quizzes are identified with their potential to help monitor and direct student learning. They are not identified, as they are in the F2F course context, as a flexible pedagogical tool which may be used collaboratively. As a result of this loss of flexibility and the dependence on CAA, the learning that results from FO quizzes has a greater association with lower-level learning than what may be expected from F2F quizzes. However, given the predominant use of MML, this may suggest participants’ views on the use of CAA are largely a reflection of their views on the use of MML.

11.6 Use of Invigilation

Invigilation refers to the use of human supervision whilst a student(s) is/are completing an assessment instrument.

Study I Background: Findings indicate that a majority of participants use invigilation. Compared to those that use invigilation, those not using any form of invigilation were found to be significantly more likely to use richer feedback – as defined by the feedback framework – and were significantly less oriented to knowledge construction – as defined by the S&B framework.

Study II - R3: How and why are participants choosing to use invigilation?

Table 32 below summarizes the use of invigilation for all study participants. In F2F courses, as accounts in this section will reflect, many of which are consistent with smaller US class sizes, invigilation is accomplished in ‘class time’ (P1) using strategies such as spreading students out so ‘they are not allowed to sit next to each other’ (P3). Typically, only summative-style assessment instruments are invigilated (as P1 discloses, this includes ‘all the classroom tests and all the quizzes’ and as P2 discloses, this includes the ‘mid-term’ and ‘final exam’). In FO courses, where there is no ‘class time’, this supervision may take place at testing centres (e.g. P2 and P5) or with a pre-approved proctor (e.g. P4).
This section is structured in four parts. First, the different factors influencing participants’ choices to use invigilation are discussed. Second, participant-identified purposes for the use of investigation are detailed. Third, identified differences in the use of invigilation between the two contexts are discussed. Fourth, some conclusions are drawn with respect to the quality of learning in the FO course context.

### 11.6.1 Identified Factors Associated with the Use of Invigilation

Participants identify two main factors influencing their choice to invigilate.

First, most participants identify a threat posed by the widespread use of cheating (e.g. ‘rampant’ as described by P1) where mathematics courses are reported to be particularly at risk:

**P4:** *I mean, I’m not naïve enough to think that students wouldn’t try to get assistance outside of class.*

**P3:** *They [students] would all cheat otherwise... they would cheat. I hate to be so cynical. They would cheat...*

**P6:** *...but more than once, we have found students listing [using Craig’s list online classifieds] that ‘Hey, I’m taking so-so’s class, so will you come and take my exams for me? I’m willing to take exams for such and such and such online classes, provide me an ID, the cost is ____.’*

While common to all academic disciplines, mathematics is identified as particularly susceptible to cheating and in need of invigilation. P4’s response, for example, suggests mathematics learning is typically demonstrated more objectively with a single answer or one of a finite number of solution paths. Whereas in other disciplines learning may be demonstrated more subjectively with writing with a seemingly infinite number of possible variations:

**P4:** *And you know with math it’s too easy to have someone help you with an answer, I wouldn’t know if they got it or not. You know, when they write a paper, maybe that’s different... It’s very*
easy in math to convince somebody that you know and you really don’t. Do you know what I mean? - you can get the answers, and I would never know who did it.

P6: ...math courses that are taught online... ALL have proctored exams – no questions. And many of our social science courses... are fully online without any proctoring.

Second, participants report various influences both internal and external to the mathematics disciplinary community – where the level of influence may be described as continuous and strong. The following section first details participant-identified sources of influence to invigilate followed by sources of influence not to invigilate. Overall (see Figure 13), the situation appears to pit those advocating for invigilation as those who prioritize outcomes such as ‘understanding’ versus those advocating against the use of invigilation as those who prioritize outcomes such as student access or enrolments.

<table>
<thead>
<tr>
<th>Influences to Invigilate</th>
<th>Influences not to Invigilate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal (e.g. mathematics</td>
<td>Internal (e.g. students)</td>
</tr>
<tr>
<td>department)</td>
<td></td>
</tr>
<tr>
<td>External (e.g. accrediting bodies)</td>
<td>External (e.g. other departments, administration)</td>
</tr>
</tbody>
</table>

Choice

Figure 13: Identified sources of influence affecting the decision to invigilate

**Influences to Invigilate**

Those most intimately aware of the nature of students and mathematics – mathematics departments composed of mathematics instructors – emerge as the chief influence in favour of invigilation:

P4: My department strongly insisted, not insisted but very strongly encouraged proctored; and I agreed, we wanted some veracity to the results that I’m obtaining.

P6: And I really appreciate the full-time faculty we have here because they’re conscientious about making sure that the material that we present is high-quality material...[Later adding, with regards to dealing with pressure against using invigilation in FO courses] all have proctored exams – no questions.

Other reasons that may influence instructors’ choice to use of invigilation are:

P2: [Helping students learn and be able to answer] simple recall questions... [and] assure[ing] that the process is fair for everybody [i.e. both F2F and FO students].
P3: I’m... so accustomed to it that it wouldn’t occur to me to chose another method.

Finally, one identified external influence relates how HE accreditation bodies, which are broadly concerned with the quality of education offered by an institution, are grappling with these issues:

P6: ...there’s a recommendation from our accrediting agency... Basically... their belief is that if you’re going to have an online course, you have to have some way of verifying the student that’s receiving credit is the student that is taking the course, etc. And so we do that verification through proctored exams.

Influences Not to Invigilate

Consistent with findings regarding the use of discussion, mathematics students, as members of the mathematics disciplinary community, are cited as an influential factor:

P1: I do not use proctoring online. I started out requiring it and then ran into situations with students who were home bound, physically handicapped, and had no contacts with anyone that would meet the criteria for proctoring and I said ‘phooey’... [In addition] I have students around the world who [sic] we can’t physically get together...

People from other disciplines outside mathematics as well as the administration emerge as somewhat strong opposing external forces:

P4: Well, that [invigilation] has been controversial here as well... the people outside the math department couldn’t understand why we wanted to have proctors...

Finally, regarding administrative pressure, P6 describes what appears to be a constant battle between the administration and the mathematics department:

P6: ...the math faculty kind of has a stain according to the, you know the administration is a little bit perturbed that all of our math courses, that are taught online or hybrid, ALL have proctored exams – no questions... the administration would like us to move to not requiring this proctoring and we, we’re just not willing to do that.

This administrative pressure may not be surprising given, as previously discussed, in the community college context where P6 teaches there is an emphasis on providing open access to higher education. In this respect, where mathematics department choose to invigilate, they may be seen to be hindering or preventing student access. As reflected in the following quote from P6 given in response to why the administration seems to be pressuring mathematics faculty not to use invigilation:
Our new motto is that we are reaching the community of students without... regard to space, time or location. In other words, we’re educating everybody no matter who they are, where they are, no matter what their capabilities are or what their location is. And so they really want everything to be fully online.

The level of influence participants identify at time appears significant. P4 describes their mathematics department as having ‘strongly insisted, not insisted but very strongly encouraged’ the use of invigilation. Similarly P6, in a related course provision issue, describes how they had to fight ‘tooth and nail’ to resist administrative pressure.

In summary, while the threat of student cheating looms large, participants report several factors influencing their choice to invigilate.

11.6.2 Purpose of Invigilation

Two purposes emerge from participants’ responses. First, the validation of student learning is identified as the fundamental emergent purpose for using invigilation. Participants’ responses identify human supervision of students completing timed assessment instruments as a necessary part of internalizing mathematics. Without this supervision, students are expected to be relying on animate and inanimate resources to complete assessment instruments resulting, for example, in ‘artificially high’ (P5) grades:

P4: [Invigilation is used because] we just want some legitimate assessment of what an individual student knows and what is able to do and put some integrity to the grade that I would be assigning them...

P5: ... [students] can use those resources to maybe get artificially high scores on examinations, on non-proctored assessments. The reality is students can go online, they can hire tutors, and they can get the problems finished... So I do think that what proctoring does is it enables you to see well what do the students really know about absent all of those resources as opposed to saying testing each student to see who is most able to use the resources that is out there. We do want to make sure that when students do their work it truly is their work not somebody else’s work.

P6: [Invigilation is a] way of verifying the student that’s receiving credit is the student that is taking the course...

Second, in contrast to the previous dominant purpose but consistent with the kind of flexibility participants may have in their assessment practice (i.e. under academic freedom), one participant
identifies what appears to be the purpose of achieving better student learning when s/he uses invigilation on half his/her exams and no invigilation on the other half.

\[ P5: \] You know historically, I have not given a lot of take-home exams, but I have to say, in the online class, I saw a lot of real-good work coming out of the students using that model; so that’s why I have indeed gone to where half of the exams are take-home and half are proctored.

In summary, with the exception of one identified perceived benefit to learning, for most participants, the purpose of invigilation is to ensure that a student grade represents actual and not some semblance of learning, which may be no learning at all.

**11.6.3 Differences in the Use of Invigilation**

Participant-identified differences in how invigilation is used in the F2F versus the FO course context are related to the required resources and the way invigilation may be carried out.

First, differences in required resources emerge as the most dominant participant-identified difference. That is, where it appears all F2F courses use invigilation and this is done in class time by the instructor, when FO courses use invigilation this means an additional commitment of time, human and physical resources. In particular, some distinction is made with regards to whether students are ‘local’ or ‘remote’. For example, more ‘flexibility’ may be required to accommodate ‘local’ student needs. However, for those that are remote, individual students are required to find an appropriate proctor. This involves getting a proctor approved by the instructor, the instructor securely sending the assessment instrument to the proctor, the proctor likely having to check ID before administering the instrument, and finally the completed assessment instrument being securely returned to the instructor. This process, when completed for each individual student, involves significantly more human resources, space and time required than F2F courses:

\[ P2: \] I need assistance for my online [i.e. FO] students... My on-ground [i.e. F2F] students, we can actually take the proctored exam all together... Because of the varied schedules of the online students... I have to be much more flexible so I will usually proctor at least two sessions... a session usually in the day time and... then our testing centre in distance education helps me out by providing an evening session for each proctored exam. If the student is not local, I simply ask them to identify a higher education institution or a military education centre where they would like to do the test and they can give me the contact information of the proctors at those locations.
P4: ...it’s a pain in the neck to – you know, we offer several proctoring times and... [students] can get their own proctor; we have to approve the proctor and we have a few proctoring sites within the...area that we permit. So... I have the proctor scan in their work and send it back to me. So it’s a lot of course management, it’s time consuming...

P5: The only difference is, in the online class the students will take the proctored exams in many different locations. They can take that in a testing lab, they can take them with a proctor out of the country, or they can take them in a classroom with me. So in the online situation, there’s [sic] many different types of proctoring; whereas in a F2F class, I am typically in there with the students.

P6: I don’t think they [i.e. the administration] want the space devoted to the testing centres because, because we have a ton of online courses... we regularly offer hundreds of online sections and if you think about how many students that is, there’s thirty students in each of those, how many testing centres and how many desks and how much employees and blah, blah, blah - there’s a real economic reason to say ‘Hey! Let’s fold up camp and not have to have these testing centers’.

A second emergent difference relates to how supervision in F2F courses is discussed in terms of everything administered in the classroom is invigilated to ‘some degree’ (i.e. formal to informal). Whereas, as previously discussed, invigilation in the FO course context is reported as a formal process (i.e. with time, place and resources strictly regulated). That is, the level of supervision in F2F courses may be represented on a continuum whereas this is better represented in FO courses as a dichotomy – either with time and resources strictly regulated or with ‘unlimited’ time and unregulated use of available resources:

P1: [Anything] done in the classroom, everything would be proctored.

P5: ...to some degree they’re [F2F quizzes] being proctored; it is significantly different in that environment than it would be in the online environment.

In summary, there is not the same expectation that invigilation will be used in the FO as in the F2F course context. As compared to its use in F2F courses, when invigilation is used in FO courses it is reported to be more formalized and require, for example, considerable more human resources.
11.6.4 Quality of Learning and the Use of Invigilation

Only two participants associate the use of invigilation directly with the quality of learning. P2 states that ‘simple recall questions’ are not ‘legitimate’ unless invigilation is used. P5, as previously discussed, reports how non-invigilated ‘take-home’ exams have led to ‘a lot of real-good’ student work:

P2: ...one of the other reasons I like the proctored exams, I want to ask my students... simple recall questions... very simple questions, and in an unproctored and open-note setting, those are not really legitimate questions.

P5: ...I saw a lot of real-good work coming out of the students...where half of the exams are take-home and half are proctored.

In summary, the issue of whether invigilation is used is a matter of real interest and concern with institutions ‘struggling in different ways’ (P6). While the use of invigilation appears to be a natural part of general F2F course assessment practice, in FO mathematics instruction its use appears to be particularly ‘controversial’ (P4), with those arguing for and against its use. As compared to the use of discussion and quizzes, which participants associate mostly with, respectively, advancing or directing learning, they associate the use of invigilation mostly in relationship to the measure of that learning. In this respect, in the FO course context, participants appear concerned that students may be using unauthorized help in completing assessment instruments and that the measure of their learning represents more an ‘appearance of rather than actual learning. Participants’ accounts suggest some tension exists between, for example, participants who want to use invigilation and institutional goals of providing access to education that may lead administrators to discourage it use. One caveat, however, is that the choices available to FO instructors may be different given, for example, while academic freedom may enable some to choose whatever action they see fit, others may be restricted by department policy. However, when FO instructors do choose to invigilate these accounts suggest this choice involves considerably more work.

11.7 Use of Feedback: Kind (and Process)

Feedback refers to information provided by participants to students about the gap between a students’ actual and some target level of mathematical understanding (e.g. Ramprasad, 1983).
**Study I Background:** Findings indicate that there appears to be no significant relationship between types of feedback used and any of the approach measures.

**Study II - R4:** How and why is feedback being provided? (Specifically, the kind of feedback used? In addition, what kind of feedback do participants identify as the most effective?)

Participants were asked to identify any differences in the kind of feedback they used in their F2F versus their FO courses. As will be shown, despite being provided with illustrative examples of what was meant by ‘kind’ (‘correct/incorrect’, ‘full solution’, ‘hints or comments’), participants repeatedly identified the question with the process of how feedback was being provided. Moreover, while no apparent differences in kind of feedback used were cited, identified process differences are seen to have an impact on the quality of the kind of feedback provided.

This section details several themes related to how feedback is provided and how this appears to be influencing the quality of learning. First, identified factors, related to how the kind (and process) of feedback may vary, are detailed. Second, emergent differences in the process of feedback are detailed. Third, participants identified what they consider to be the ‘most effective’ kinds of feedback. Fourth, conclusions are drawn with respect to the quality of learning in FO courses. Findings mirror those regarding the use of quizzes and discussion where issues associated with the use of CAA and the loss of F2F contact suggest limitations to the potential quality of learning in FO courses. Specifically, while participants may be using the same kind of feedback in both contexts, process differences in how that feedback is delivered appears to diminish its overall quality and thus suggest a potentially poorer learning experience.

**11.7.1 Identified Factors Associated with the Kind (and Process) of Feedback**

Participants identify several factors related to the kind (and process) of feedback. The first three factors are student demographic characteristics, type of assessment question asked or instrument used and the context of the assessment. The fourth factor is feedback agency where, consistent with findings regarding the use of quizzes, computer agency figures significantly.

**Student Demographic Characteristics**

Participants identify student demographic characteristics as a factor influencing the kind (and process) of feedback:
P3: Yes it’s very different [F2F vs. FO]. The feedback I give my online students is much more detailed. They deserve it - they’re math majors.

P5: ...the types of feedback I give are probably not as detailed as they are in the online class where I think the students are more independent...

**Type of assessment question or instrument**

The type of assessment question (e.g. solving a problem versus providing a definition of a mathematical term) or, as may be expected, instrument used (e.g. exam versus homework) is seen as an influential factor:

P4: Well, I guess it depends first of all not only on the feedback but on the type of assessment that you give: do I ask the students to explain their thinking about something and how they got to something and what is their understanding about it. And then if I would provide detailed feedback to that, you know, like you say, there’s two different kinds!

P5: ...[when] students are working on a quiz and they are having difficulty, I can give the students in the F2F class hints... I don’t give a lot of feedback on the homework because I don’t collect it very much, but what I do is I provide complete solutions...on a final exam, you won’t give a whole lot of feedback because the students don’t receive them; I record them, I give them a grade for the class. Now for a midterm exam, I’ll give them some feedback, I’ll explain to them where they did things incorrectly and they can learn from that.

P6: In my F2F classes, I use quizzes because I don’t grade the homework that is turned in in my F2F classes. Homework... you get some points because you’ve turned it in... The quizzes are a chance for me to actually say ‘You either understood it or you didn’t’...

**Feedback Context**

The context in which feedback is provided also appears to be linked to differences in feedback provision. For example, submitted assessment instruments receive feedback directly on the assessment instrument (written), in-class (verbal) and/or on an online discussion board (text-based). In general, these differences appear related to whether the feedback is given in the F2F or the FO context (as per responses from P5 and P6 below). However, as the response from P2 reveals, the kind (and process) of feedback may also vary within a single instrument based on whether the problem is being attempted, for example, for the first or second time:
P2: ...they have two attempts on chapter quizzes and after their first attempt they don’t receive any information about what they missed - they don’t know which questions they’ve missed... after the second attempt, I will give them some feedback on what they got wrong and what they might want to review or study...

P5: Well, with respect to written work, so these are our exams... I typically provide solutions to the online class, so I will pick up solutions or I’ll do a pen cast with the solutions... Whereas in a F2F class, I’m more likely to give them oral feedback when I pass back examinations and assessments.

P6: ...my first ten minutes of class – of every class no matter what I teach in F2F – is: ‘What are your homework questions?’ And I answer them. And in a fully online course, we have this homework discussion and I make sure that either I answer or students answer every homework question.

Feedback Agency: Computer

The use of CAA and computer feedback is perhaps the most significant factor. Though limited by the predominant use of MML, analyses of participants’ responses reveal at least two emerging characteristic ways CAA-based computer feedback is associated with the kind (and process) of feedback.

First, evidence from participants only using MML suggests that while the feedback is individualized it focuses only on the answer that is inputted and not the underlying mathematical thinking. Evidence of this is reflected in the ‘point’-based nature of the feedback where the points are associated with whether an answer is right, wrong, or partially right. Alongside this, any additional feedback is somewhat generic with help offered in the form of linked resources that address the mathematics underlying an individual problem or how a similar problem may be solved. In short, while this may be primarily a characteristic of MML, the associated feedback does not appear to directly address an individual student’s specific underlying thinking:

P1: Homework online [i.e. CAA] they have access to everything that MyMathLab - ‘help me solve this’, ‘show me an example’... when it comes to exams online, the students can see nothing except their score until the last student has finished the material in that particular exam and then they can go over it...

P2: But every [CAA] problem they work, they get individual feedback on that problem. For their quizzes... after their first attempt they don’t receive any information about what they missed – they... make their best guess where their weakness is, and they get to re-attempt that quiz, after
the second attempt, I will give them some feedback on what they got wrong and what they might want to review or study so it’s not missed the questions on their exams.

P4: [With CAA] I don’t see their individual work but I know what mistake they made if they got a specific answer... [Later adding]... it doesn’t always explain why they got it wrong. You know we were working on our home-grown software, where we tried to have the computer diagnose what the students, you know if they made this mistake that means they get this, and we never got that far with that. You know, that’s difficult to do on a computer I think.

Second, in what may be argued to be a related phenomenon, participants identify a propensity for students to focus on question and answer patterns and not the underlying mathematics that yield correct answers. Indeed, evidence of this emphasis comes from participants’ accounts where the CAA feedback appears to be conditioning some students to obtain the ‘correct’ answer without necessarily having ‘correct’ mathematical understanding. Both P5 and P4 refer to student ‘pattern’ seeking:

P5: ...they can do is memorize the patterns of the feedback and they can go ahead and give that back to the computer and answer right without really knowing why something is working.

P4: Well let me tell you one bad experience...so we teach them [students] how to solve a cubic or a quadric using synthetic division, the rational and zero theorem and all that.... so I had this student come to my office and say, ‘Oh, I know how to solve those!’ And I said, ‘How do you do it?’ (This is an online student), he said, ‘Well, the answer is always the constant over the leading coefficient.’ He didn’t call it that, he pointed to the constant over the leading coefficient, not the factor of the constant over the factor of the leading coefficient; ... ‘And the other two answers are always one plus or minus the square root of two.’ I said: ‘Well how do you know that?’ ‘That’s how it always works out on MML [i.e. MyMathLab].’ And here what I had done was, they didn’t have enough of this one type of problem so, I naïvely, put in the same problem from the book - maybe three or four times - and, structurally worked out to be the constant over the leading coefficient plus or minus the square root of two. They didn’t really change the problem when they algorithmically generate it, it’s structurally the same - so the student could see that pattern and they could get the answer right, without knowing the mathematics, which is kind of disturbing!

In summary, participants’ responses reveal a number of factors influencing the kind (and process) of feedback. These factors suggest inherent complexities involved in investigating the kind (and process) of feedback provision in both F2F and FO courses.
11.7.2 Differences in the Process of Feedback Used

As mentioned, participants repeatedly identified the interview question with the *process* of feedback provision. No immediate differences in the *kind* of feedback are identified. As a result, six process-related differences are identified. First, the amount of feedback provided appears to emerge as a primary and significant difference. Second, differences are identified with respect to whether feedback is provided by question or by assessment instrument used. Third, differences are reported in the way feedback can be supplied to the entire class. Fourth, differences related to how feedback is targeted to individual students are discussed. Fifth, differences in required time and effort needed to provide feedback are detailed. Sixth, differences in emphasis on the use of office hours are discussed. The main emergent theme is that, with the loss of live F2F contact, the processes currently used to provide feedback in FO courses are producing a qualitatively inferior kind of feedback compared to what may be experienced F2F.

*Differences in Amount of Feedback Provided*

When asked about the kind of feedback used, participants seem to instinctively equate ‘kind’ with ‘amount’ of feedback. And while most participants state that they give more feedback in their FO than their F2F courses, these differences appear complex.

As an example of what appears to be an instinctive association of ‘kind’ with ‘amount’, two participants, directly after being questioned, make immediate reference to the amount of feedback:

*P6:* I would like to say that the students in both courses get equal kinds of feedback - [I: Kind not amount?] Okay, then in that case, the kind of feedback, for the most part, is the same...

*P2:* [I: What I’m hearing you say is that in the online setting, the amount of feedback you give is greater then what you give in the F2F; but in terms of kind of feedback, in both settings, you get basically the same kind?] That is right.

In a similar fashion other responses, some mentioned earlier, refer to the amount of feedback provided and reflect the complex nature of feedback processes. In particular, P4, who uses MML and P5, who makes no reference to MML, contrast differences with regards to whether CAA is used with exams. That is, with the use of CAA and computer feedback, P4 states that his/her FO students get less feedback than his/her F2F students. Conversely, not using any CAA
but providing both ‘written’ and ‘oral’ feedback, P5 describes the reverse effect with regards to his/her ‘written’ feedback:

*P1:* Yes...in some sense, the students online get much more feedback regarding homework and much less feedback regarding exams than students on campus [i.e. F2F].

*P4:* ... [For FO] homework feedback; I will often in MyMathLab... I could go and see who’s completed certain assignments and come up with an email for anyone that didn’t complete this assignment... and then I would give them more detailed feedback than I would normally give them in a F2F with homework... [However, regarding FO exam feedback, s/he says] I don’t think they get the feedback as detailed as they do in a F2F class.

*P5:* ...in an online class, I’m much more likely to give detailed feedback because the students do not have me in person for that feedback so when I return an exam to the students, I’m going to be more deliberate with my notes, I will give them more feedback, more written feedback. Whereas in a F2F class, I’m more likely to give them oral feedback when I pass back examinations and assessments.

In summary, participants emphasize differences in the amount of feedback provided when asked about differences in the kind of feedback provided. While differences in the amount appear complex, participants’ responses suggest that, overall, they provide more feedback in their FO than their F2F courses.

**Differences in the Provision of Feedback: By Question or Instrument**

One participant identifies a difference related to whether feedback is provided after a question is completed versus after an entire assessment instrument. This difference appears related to the dependency on MML in their FO course, where computer feedback is associated with feedback provision per individual question whereas instructor-provided feedback is associated with feedback provision for an entire assessment instrument:

*P1:* F2F we go over homework approximately once a week so the students have to complete assignments on their own... feedback from the instructor would occur when we go over that which is presumably completed. Homework online they have access to everything that MyMathLab... there definitely is a difference about the way it’s handled. The students online tend to get their feedback much more problem-by-problem, and students on campus tend to get their feed back in a chunk after they’ve presumably completed the assignment.
Differences in the Provision of Feedback to the Entire Class: In-Class F2F vs. Class-Wide FO

Participants further identify differences in how feedback is provided to their entire class. While every participant acknowledges the use of F2F class time for in-class feedback, and while participants’ responses suggest they are attempting to offer similar feedback processes (i.e. covering assessment questions with instructor or peer feedback accessible to the whole class) in their FO courses, the effectiveness of this process is seen, on balance, as inferior to what may be accomplished in the F2F context. In particular, when compared to in-class F2F feedback, some of the key emergent differences relate to the optional nature of class-wide FO feedback as well as the asynchronous nature of interactivity in FO courses.

As discussed previously regarding the use of discussion, P5 contrasts the potential of the F2F classroom environment – where ‘informal’ feedback can be offered – with the lack thereof in the FO course environment:

P5: Again, I think any sort of feedback from me is more important in the online class because I can give them informal feedback in a F2F class. If I have a class that meets four days a week and it’ll take me a week to grade the exams; on the day of the exam I’ll talk about the difficult questions. If the students have questions about the exam, I can go over the exam in that environment, we have more informal discussions. Whereas in the online class, that’s not really possible... And as far as the non-proctored exams, the take-home exams, again, I don’t give a lot of feedback to the students on those in the online class - though I have to say in the F2F class, students will occasionally ask me questions about those things as well...

However, in FO courses, some attempts are being made to offer feedback processes similar to what may be provided thorough in-class F2F feedback.

First, P4 and P5, for example, state how they use screen recording technology as means of providing class-wide feedback in their FO courses:

P4: But... if I found students emailing me, a lot of students emailing me the same question. I would take a Camtasia video with my solution - talking through it, working it out; so I try to do it that way.

P5: I typically provide solutions to the online class, so I will pick up solutions or I’ll do a pen cast with the solutions and I’ll provide those to the online class. In the F2F class, I go over many of the questions on the exam...
Second, the use of online discussion emerges as another potential means of providing class-wide feedback in FO courses. Consistent with earlier discussed findings on the use of discussion as a weighted assessment component, participants do not consider the quality of this feedback to be as good as what may be provided as in-class feedback in their F2F courses:

P2: The [FO] discussion board... can help fill the gap a little bit, but if there’s a real issue or concept of something that needs some attention, I just want the student to have attention very quickly.

P6: ...sometimes I put feedback on Blackboard [the course VLE], which has a little space where I can type in the feedback to the assignment. Most often they don’t ever respond to that... They kind of blow you off [ignore you] a little bit more than they would in a F2F... [And later s/he adds:] But the difference is that in the, in the online course they watch the video on their own time and they write a response to those discussion questions; whereas, in my classroom, they watch the video and then we dialogue about it in the classroom. And so that dialogue doesn’t always happen when they’re just sending a written assignment and sending it in and sending it in, even if I give some feedback...

In comparing the use of screen recordings and discussion boards to in-class feedback, the optional nature of engaging with feedback also appears to surface. That is, F2F in-class feedback is at least witnessed, in some way, by all students attending a class. However, FO class-wide feedback appears to be completely optional. As in the previous responses from P2 and P6, the following response from P1 illustrates the optional nature of engaging in some FO feedback:

P1: Yes. When it comes to exams in class [i.e. F2F], we go over them, when it comes to exams online... I don’t go through the exam or make any attempt to, but if they ask any questions about it, I respond to those.

Finally, as with previously discussed expectations concerning the use of discussion, both community college participants identify one further difference which contrasts the potential of synchronously interactive F2F in-class feedback with the limitations of the typically asynchronously interactive FO class-wide feedback:

P5: When students are collaborating [F2F], I can check to see what kind of work their doing, I can give them hints, I can give them suggestions; I can do some scaffolding, I can ask them questions that help them see the connections. So again it’s a more interactive model, the feedback that I give them is immediate and it’s interactive. Whereas in the online class, if they
are working... on a collaborative quiz offsite which they have kind of given up, there is really no way for me to do that... So whenever you’re doing feedback, the feedback should go both ways, I guess. So I talk to a student, the student talks to me, we engage in a conversation, and from that conversation, I think students can develop understanding.

P6: ...my first ten minutes of class - of every class no matter what I teach in F2F - is what are your homework questions? And I answer them. And in a fully online course, we have this homework discussion and I make sure that either I answer or students answer every homework question; but there’s something different about me typing in a discussion... when I have a student in a classroom, and they ask the homework and I’ll say come put it up on the board and show me what you’ve done and there’s something about the, you know, the asynchronous dialogue [i.e. in the FO context]... it’s lacking... that synchronous give and take that happens in the classroom...
So I would love to say that the KINDS are the same but I think that asynchronous vs. synchronous make those kinds different, even though I’d like to say that they’re the same kind...
[Then when asked to clarify the differences in the process and/or the kind of feedback responds]
Okay, or maybe, maybe I’m talking about the qualitative value of the feedback... because I think that in the interaction that happens when the student and I are standing at the white board together, working on the problem together, it’s qualitatively different than when a student and I are dialoguing in an asynchronous fashion.

As stated in the last quote, while the same kind of feedback may be used in both contexts, it is the differences in process that appears to change the ‘qualitative value’ (P6) of that feedback. In particular, while screen casts and online discussion boards offer some value, on balance this value is seen to be largely inferior to what may be provided F2F in-class given, for example, the asynchronous nature of FO courses and the comparatively voluntary nature of engaging in any class-wide feedback. Again, such responses from these two participants may be expected given community colleges typically have smaller classes and a greater focus on teaching.

**Differences in the Provision of Feedback to Individual Students**

Possibly in an attempt to mediate for the loss of the F2F in-class feedback experience, some participants are providing more individual feedback to students in their FO courses. A shift that seems consistent with participants’ reports of more feedback being provided in FO courses:

*P3: Yes, I do this on an individual student basis. It takes me about 2 ½ to 3 hours – I generally have 12-15 students in my course. It takes me about 2 hours to grade a problem set per student...*
[In contrast, F2F students\(^{35}\)] don’t get that much feedback on their graded exams. They get comments like ‘wrong’ and they get partial credit; but they don’t get extensive comments on what was wrong. But I do, after I’ve given the exam, I do with the next class say... ‘does anyone have questions on their exams?’

P5: ... here is a big difference [between F2F and FO feedback]... when I return an exam to the [i.e. FO] students, I’m going to be more deliberate with my notes, I will give them more feedback, more written feedback. Whereas in a F2F class, I’m more likely to give them oral feedback when I pass back examinations and assessments.

**Differences in Required Time and Effort**

Both community college participants also note the extra time and effort required providing feedback in their FO courses. While this is consistent with the previously discussed provision of more individualized feedback, it is also consistent with expectations that these participants are attempting to provide similar feedback in their FO courses as they are accustomed to provide in their F2F courses:

P6: But if they’re lost on a whole assignment or if they don’t understand an entire module... giving that kind of detailed feedback gets overwhelming when I’m teaching two online courses with this much work and sixty students... [Then sharing how much feedback s/he has provided in the first nine days of her FO course] We’ve posted 260 discussion messages; 37 of which I’ve posted, so, I’m posting a little more than 10% of the time. [I: Is that a lot of work?] It is...

P5: ... if it can be F2F, and in real time, I think we can make a lot more progress. Now I do that sort of thing in the online class, but it’s usually asynchronously, so I’ll ask a question, the student will respond an hour later and then we’ll go from there, so I don’t think that’s nearly as effective as it is an a F2F environment.

Agreeing with these responses, P4 provides some further explanation as to why these differences may exist. S/he sees at least two stages in the feedback process, both considered more difficult to undertake in the FO course context. First, the ability to ‘evaluate’ student ‘thinking’ and, second, the ability to provide ‘detailed feedback’ that helps the student to ‘grow beyond’ their current state. Additionally, it is further noted that any efforts to try explore student thinking

\(^{35}\) It should be noted, as previously discussed, both the nature of the courses and the students taking them are different. That is, the former referenced course is on the use of Mathematica and is taken by mathematics majors and the latter is a Calculus course taken by non-majors.
through the use of CAA (specifically MML) is ‘a whole lot more time-consuming’ given the necessity of reviewing and then writing a response to individual problems. This is expected to be a more significant issue for FO courses given the assumed greater dependency on CAA:

*P4:* ...so there’s definitely pros, and in terms of assessment... on the Vista assignments [those uploaded and accessed through the VLE] I probably actually responded in more depth online because I would download their Word documents and make individual comments probably more so than when I grade a similar assignment by hand - when I would go through it kind of quickly. So maybe, for deep learning, on anything except the MML assignments, there is potential for better assessment and deeper, I think. But it takes effort and time... I want to encourage them to look at what their thinking and help them grow beyond that and then I also have to evaluate where they’re at. So, I guess in the F2F, it’s easier to do both. In the online [FO course] it’s more difficult to do both. [I: How, when using CAA in either context, do you provide detailed feedback on students thinking?] Well, you know it’s harder but you know, it’s more time-consuming... on MyMathLab... when they do a test and I review it... I don’t see their individual work but I know what mistake they made if they got a specific answer. So I... sometimes go... and look at the individual problem and write a response to it... but it’s a whole lot more time-consuming.

**Difference in Emphasis on Office Hours**

Perhaps, in part, due to the extra time and effort required to provide effective feedback in the FO context, some participants are encouraging students to come to their office hours to meet F2F:

*P6:* ‘Please come meet me in my office hours, let’s have F2F dialogue!’... some students do that and when they do some say ’Wow, I’m going to come see you every week.’ I say ’Yes, good plan. We could almost turn this into a F2F class’.

*P4:* ... I had online office hours... [and] I had F2F office hours, for [FO] students who were on campus. And honestly, like, I would have students that would spend a lot of time with me... So this one handful of students got a ton of individual attention. But the whole class didn’t and that’s a big difference... But two or three students do that and they get wonderful attention and probably more detailed feedback then they would in a F2F class...

36 F2F office hours, already available to F2F students, may be used by FO students if, for example, they are local students who live close to the institution. In at least one case, P3 schedules F2F office hours specifically for FO students:

*P3:* With my online [i.e. FO] students... I schedule my office hours by appointment. For my online students, I tend to schedule office hours on Friday evenings from 5-7.
P3: ... in [F2F] mathematics courses they get F2F feedback. In the Mathematica course [FO] they get written feedback plus an encouragement to see me F2F.

This last statement, almost amounting to an equation, presents what most if not all participants seem to express concerning the use of feedback (and discussion) in the FO mathematics course context. That is, the loss of live F2F interactions in FO courses, which participants link with effective assessment practice and instruction, is not currently being bridged. Instead, for example, students are being ‘encouraged’ to seek out a live F2F presence through the use of office hours.

In summary, there appears to be no immediate apparent difference in the kinds of feedback used. Instead, participants’ responses seem to elicit only accounts of how the process of feedback differs between the two contexts. These identified differences in process point to, for example, more feedback which is more individualized in FO courses with participants spending more time and effort providing FO feedback. While class-wide feedback is generally used in both contexts, participants identify class-wide FO course feedback with a loss that corresponds to common attributes of F2F class-wide feedback (e.g. real-time interaction). This loss appears evidenced in the emphasis on the use of office hours in FO courses. In conclusion, while on the surface the same kinds of feedback are used in both contexts, participants’ responses suggest the nature of current FO courses, and the processes they allow, is rendering the quality of this feedback inferior to what is experienced F2F.

11.7.3 Identified ‘Most Effective’ Kind of Feedback

As part of the next question (i.e. on the purpose of feedback), participants were asked what kind of feedback they considered to be the most effective ‘in helping students understand mathematics’. These findings present additional evidence that the majority of participants emphasize the process of feedback when asked to consider the kind they use. Additionally, findings for this question suggest participants view how feedback is provided as more important than what that feedback actually is. This section first discusses those feedback processes identified as most effective. Following this, the identified most effective kinds of feedback are discussed.

First, participants view feedback that is ‘one-on-one’ or ‘side-by-side’ as the most effective. As with the use of discussion, the emphasis appears to be on direct F2F interaction that is, for example, ‘personal’ and in ‘real time’:
**P1**: I think one-on-one feedback is probably most effective because... one-on-one feedback [is] in the immediate situation... just in time... [and] personal. When we do feedback to the whole class there are, of course, those who tune out because they don’t need it, or they think they don’t need it, or there is something else that’s more interesting... Ideal feedback to me would be to work side-by-side with the student who is having difficulty, guiding them, as they are attempting to do the work.

**P5**: I think that the most effective feedback is in person feedback, where I can sit down with the student and I can talk to that student about the types of errors s/he made, the types of thinking s/he has... So, if it can be personal, if it can be F2F, and in real time, I think we can make a lot more progress.

Second, as what may be inferred from the previous emphasis on ‘one-on-one’ contact, there is also an emphasis on interactivity. Consistent with previously discussed expectations, this emphasis comes from both community college participants. Both of the following quotes refer to a feedback process that is a two-way exchange:

**P6**: I’m looking for an ‘Aha!’ moment to happen and I feel like those happen most often in a give and take in a discussion.

**P5**: So again it’s a more interactive model, the feedback that I give them is immediate and it’s interactive...And that we can engage in a conversation that I can attempt to understand how the student perceives the material, so I can see the gaps in the understanding, I can see the places where s/he is failing to make the connections; and then I can do some scaffolding to help that student get to where they need to be...

Third, as stated in most of the previous quotes, participants also identify feedback that is ‘immediate’ as a significant characteristic of the most effective feedback. Suggesting a desire for live interactivity, both P1 and P5 use the word ‘immediate’ while P2 refers to feedback that can be received ‘immediately’.

Participants’ views about the most effective kinds of feedback, expressed both indirectly and directly, are limited. For example, in what may be considered indirect, P2 talks about ‘point[ing] students in the right direction... [so that] they have to dig’. However it is unclear what kind of feedback is used to ‘point’. Similarly while P5 emphasizes engaging students ‘in a conversation... [and doing] some scaffolding to help that student get to where they need to be’ it is again unclear what kind of feedback is used to engage in that process.
Most references to kinds of feedback are direct. And these responses identify ‘correct’ thinking (e.g. right/wrong, full solution) as being the least effective and feedback that provides hints, comments or questions student thinking as being most effective:

\[ P1: \] I don’t want to do the work for them. I want to lead them towards what they need to do, help them strain out what’s extraneous in their mind, make suggestions... ‘now you might do this but what about?’ – those kind of questions.

\[ P3: \] I find hints and comments are much better than right and wrong. I mean I give hints and comments then I just give a total score.

\[ P6: \] I think it’s the challenging questions that have been in the discussion; where I say ‘Can you explain that?’

In summary, while some kind(s) of feedback are identified, consistent with the findings on differences, views typically focus on the process of feedback provision. When these views are described (feedback that is one-on-one, interactive and immediate) and taken together they resemble what may be expected in F2F course feedback. This is consistent with previous findings suggesting a feedback experience that is both lacking and sought after in current FO courses.

11.7.4 Quality of learning and the Kind of Feedback Provided

The ‘kinds’ of feedback used in either context appear to be the same. For some participants this feedback is identified as qualitatively better in their FO than their F2F courses. For example, it is reported to be more ‘detailed’ (P4 and P5), provided more question by question (i.e. versus instrument by instrument; P1), more individualized (P3 and P5), as well as recorded for possible review (P4 and P5). However, despite these apparent plusses, on balance, participants identify process differences that render the overall quality of feedback provided in FO courses inferior to that provided in F2F courses.

These differences may be organized as those related to, what will be termed, ‘initial feedback’ provision and those related to what will be termed ‘feedback interactivity’ that, for example, follows from initial feedback. That is, initial feedback is the feedback given as part of grading a question or an instrument (e.g. as part of a returned graded exam) and feedback interactivity is the two-way interaction that, for example, proceeds from the provision of the initial feedback (e.g. a student asks the instructor to explain how to solve a problem they got wrong). Using this
lens the qualitative differences between F2F and FO feedback provision, particularly regarding feedback interactivity, become clearer:

First, while initial FO course feedback provision may be qualitatively better, these differences reflect on the nature of feedback agency. That is, for example, as P4, using MML, states, if the feedback is provided through human agency it may lead to ‘deep learning’. However, this is not seen to be the case, as further supported by the findings on the use of quizzes, if the feedback is provided through computer agency:

\[P4: \ldots I \text{ probably actually responded in more depth online because I would download their Word documents and make individual comments probably more so than when I grade a similar assignment by hand – when I would go through it kind of quickly. So maybe, for deep learning, on anything except the MML assignments, there is potential for better assessment and deeper...}\]

\[P5: \ldots \text{here is a big difference: in an online class, I’m much more likely to give detailed feedback... so when I return an exam to the students, I’m going to be more deliberate with my notes, I will give them more feedback, more written feedback.}\]

Still, the status of any claim or evidence of qualitatively better initial feedback in FO courses remains unclear. Of the five participants teaching identical F2F and FO courses, responses suggest only two (i.e. P4 and P5) are providing qualitatively better initial feedback in their FO versus their F2F course. Furthermore, as supported by the study findings on the use of CAA-based quizzes and limited by the predominant use of MML, initial feedback provided via computer agency (e.g. as exemplified by P1, using MML, whose question by question initial feedback was via computer agency) is considered as qualitatively inferior.

Second, the quality of FO course feedback interactivity is identified as qualitatively inferior to what may be provided in the F2F course context. Many of the reasons for these differences follow from those cited with regards to the use of discussion (e.g. unnatural, formal, reduction in multimodality, ability to judge and execute timing...) and reflect on the asynchronous nature of FO courses that shape the nature of these interactions. As previously discussed, community college participants (i.e. P5 and P6), who may be more sensitive to changes in the teaching environment, state:

\[P5: \text{[Referring to the most effective feedback]} \ldots \text{if it can be personal, if it can be F2F, and in real time, I think we can make a lot more progress. Now I do that sort of thing in the online class, but it’s usually asynchronously, so I’ll ask a question, the student will respond an hour}\]
later and then we’ll go from there, so I don’t think that’s nearly as effective as it is an a F2F environment.

P6: So I would love to say that the KINDS [of feedback] are the same but I think that asynchronous vs. synchronous make those kinds different, even though I’d like to say that they’re the same kind...

Additionally, contrasting with students attending F2F classes who are either silent witnesses or active participants, three participants (P1, P2 and P6) identify student engagement in FO course feedback interactivity as optional. Perhaps most importantly, while F2F students have the choice to ‘tune out’ (P1) any feedback interactivity, FO students must first choose whether to ‘tune in’.

In effect with the use of feedback, as with the related use of discussion, it appears that the loss of direct F2F contact is presenting a difficult learning environment where making ‘progress’ is ‘a lot’ (P5) more challenging. In balance, compared to F2F, participants experience the FO course context as a much more challenging environment for realizing quality student learning through the use of feedback.

11.8 Use of Feedback: Purpose

Study I Background: Findings suggest that there is no significant relationship between the kind of feedback instructors use and their approaches to teaching and assessment – as measured by the study measures.

Study II - R4: How and why is feedback being provided? (Specifically, the purpose feedback is used?)

Participants were asked whether the purpose of feedback was viewed any differently when teaching their F2F and FO courses. Responses reveal five purposes and while no differences emerge between contexts in terms of the overall purpose, some participants see differences in terms of intermediate (P6) purposes.

This section first begins by detailing the five purposes. Second, identified differences in purpose are detailed. Third, the section concludes by considering the quality of learning in the FO context as related to these differences in purpose. Similar to previous findings, the analysis provides yet another lens through which the FO course context continues to be seen as a challenging context for realizing high quality learning.
11.8.1 Purpose of Using Feedback

Participants identify one principal purpose for the use of feedback and four intermediate purposes that occur along with and within this principal purpose. First, the principal emergent purpose is to facilitate the forward movement of student understanding towards the goal of achieving what participants state as, for example, ‘competency’ (P1) or ‘success’ (P5). Second, feedback is identified for the purpose of helping students develop their understanding. Third, feedback is identified for the purpose of identification of wrong answers and/or thinking. Fourth and fifth, respectively, the remaining two purposes are that of encouraging self-assessment and interaction.

Feedback to Move Forward

Every participant appears to identify the purpose of feedback with and express it as some form forward movement:

P1: I want to lead them towards what they need to do.

P2: ...point them [students] in the right direction.

P3: ...go further with the material.

P4: ...grow beyond [their current] thinking.

P5: ...improve performance, so we can move them forward... [Later also referring to using feedback as part of getting students] to where they need to be.

P6: ...push them forward to go back and re-think whether or not they’ve gained that understanding.

Each of the remaining emergent purposes, which will now be discussed, fall under this principal purpose.

Feedback to Gain Understanding

Every participant except P1 appears to broadly identify the goal or purpose of this forward movement as gaining ‘understanding’.

P3: ...feedback has helped their understanding and they can actually get it right.

P4: [Explains his/her emphasis on monitoring and developing student] thinking... [and]... understanding.
P5: [Talks about] develop[ing] understanding.

P6: I mean ultimately, I want the students to understand the material and the reason for feedback is to try and either acknowledge that they have gained that understanding or push them forward to go back and re-think whether or not they’ve gained that understanding.

The specific nature of this understanding remains somewhat unclear. Two participants, referring to ‘connections’ and ‘concepts’, provide some clues:

P2: My purpose in feedback is in the direction in helping the student to understand that particular problem or concept.

P5: I can see the gaps in the understanding, I can see the places where s/he is failing to make the connections...

In summary, while the development of student understanding appears to be the goal of forward movement, the nature of that understanding is largely unclear.

Feedback and the Identification and Utility of Errors

As identified with the most effective kind of feedback, in order for students to develop understanding, the role of errors is identified as playing a pivotal role. P6, for example, discusses movement that resembles the feedback loop, discussed earlier (Hounsell et al., 2008), where forward movement may involve a ‘step back’ that is triggered by getting a wrong answer. As such errors are seen as helping students first identify and then also ‘learn’ (P4 and P5) from their wrong answers and/or thinking. Several participants’ accounts refer to one or both of these purposes:

P1: In all cases, the purpose of the feedback is to make sure the students have an opportunity to see what is correct... [so as to] help... strain out what’s extraneous in their mind...

P3: ...lots and lots of feedback about what they’ve done wrong; with the hopes, that when they finally submit it for final grading... they can actually get it right.

P4: [Refers to using feedback to help students] learn from their error...

P5: ...if they made errors, I want them to find those errors, I want them to learn from those errors...

In short, errors seem to be valued for their role in facilitating students’ forward movement, redirecting their thinking and advancing their learning.
Student Self-assessment

Along with the identification of errors, the process of who identifies them emerges as another purpose for the use of feedback. Specifically, in the following quotes, all embedded in previous responses, some participants identify the use of feedback with student self-assessment:

P4: I want to encourage them to look at what their thinking and help them grow beyond that...

P5: ...I want them to find those errors...

P6: I kind of need the students to recognize they don’t understand something.

As part of the feedback process, these statements seem to emphasize the need for students, rather than CAA or the instructor, to recognize their own errors. As P5 later states, s/he links the ability to self-assess to building the ability of students to think meta-cognitively and to be ‘self-confident’.

Interaction

A final emergent purpose, also serving the forward movement of student understanding, is interaction with the mathematics through the use of course material and/or with peers and/or the instructor. Participants identify a two-fold purpose related to interaction with the first purpose to stimulate and the second to maintain interaction:

P3: I give feedback to encourage them to learn the stuff.

P4: I want to encourage them to look at what they’re thinking...

P5: [Discusses using feedback] to provoke the students, to get them to engage with the material. [S/he goes on to separate ‘academic’ from ‘behavioural’ feedback – where ‘academic feedback’ appears to directly refer to feedback on the mathematics and ‘behavioural feedback’ appears to refer to, for example, student engagement.]

P6: ...sometimes I give feedback, but I don’t feel like it... generates the same sort of discussion...

In summary, the predominant view is that feedback is a means of advancing student understanding of mathematics. As identified by participants, this is accomplished when feedback is used to prime engagement, identify erroneous answers or thinking, help students learn how to assess their own work and help stimulate and maintain interactions. The following response brings together many of these purposes:
P5: I can talk to that student about the types of errors s/he made, the types of thinking s/he has. And that we can engage in a conversation that I can attempt to understand how the student perceives the material, so I can see the gaps in the understanding, I can see the places where s/he is failing to make the connections; and then I can do some scaffolding to help that student get to where they need to be.

11.8.2 Differences in the Purpose of Feedback

Participants appear evenly divided about whether they view any differences in the purpose of feedback between the two course contexts. For example, P1, P2 and P3 view the ‘intent’ (P1) or ‘end-point’ (P3) as essentially the same but P4, P5 and P6 recognize some differences. The apparent disagreement between these two groups may be understood by considering what P6 refers to as ‘ultimate’ versus ‘intermediate’ or ‘consequential’ purposes. That is, all participants contend that they use feedback for the same purpose, which is to move students’ understanding forward, however, some recognize different ways in which feedback is being used in reaching that end.

Three differences are identified. For the first two, when compared to F2F courses, FO course feedback is identified as being used more for helping students self-assess and more likely to be used for the purpose of stimulating and maintaining student engagement and interactions. The third difference, identified by one participant as more of an outcome of the greater dependency on CAA and computer feedback in FO courses, is that this feedback is ‘more summative’.

First, introducing the idea of intermediate feedback purposes, P6 states how s/he is able to naturally assess students in the F2F course context by reading ‘their faces’. This appears consistent with expectations that, in the absence of being able to physically observe their students, there is a greater necessity for students to be able to self-assess and inform the instructor if they are having any difficulty:

P6: ...there’s a sort of [sic] an intermediate goal in the online class. I kind of need the students to recognize they don’t understand something. Whereas I feel like in the, in the F2F classroom, that’s really easy for that to happen because I could see their faces and they can see mine and when I see that ‘deer in the headlight’ look, like ‘oh, they’re totally lost’.

A second intermediate purpose is that FO course feedback appears to be used more for the purpose of student engagement with the material and with each other:
P5: ... one concern I have is that those students do not engage material as much as they should... I have to constantly monitor the participation rate of the online students, I have to consistently send emails, phone calls - whatever I’m going to do - to provoke the students, to get them to engage with the material. If students are falling behind in the online class, I’m much more likely to give them feedback right away... Whereas, I don’t need to do that as much with the F2F classes. I mean if I have students who are not attending class... I don’t lean on them nearly as much as I do in terms of leaning on the students who don’t complete their assignments online. So yes, I absolutely give consistent feedback to assure the participation of the students in the online class.

P2 describes a similar proactive stance whereby the purpose of the feedback in FO courses is to reach out and prompt the student to engage with the instructor or the material. While in the F2F context, any instructor-student or student-content engagement is left to the student to initiate:

P2: Now if a [F2F] student comes to my office they will receive a little bit more feedback, but the student has to self-select to do that, I can’t make them come to my office. Whereas in the online course, I just go to the student, I don’t wait for them to ask.

Third, one participant appears to link the F2F course context to the potential for ‘more formative assessment’ and the lack of a similar context, with a regular meeting time and place, with feedback that is ‘more summative’. Though using MML, s/he also links the greater dependence on CAA feedback in FO courses to feedback that is ‘more summative’. His/her response is consistent with what was previously discussed regarding the more optional nature of engaging in FO course feedback and seems to imply this feedback will remain summative unless a student initiates a F2F meeting:

P4: ...in the F2F, I can give more formative assessment, which means you know I can help them, I can see what they are doing, what their level of thinking is, and encourage them to go one step beyond. Where in the online, it’s more or less... you take the exam if you want to come in I’ll go over it with you... it’s more summative in the online... [I: Does this have anything to do with the fact that this feedback is computer-generated?] Yes, I do... If I would have them do homework and scan them in to me, then you know from a book as opposed to an online, algorithmically-generated problem homework management system. It could be different...

In sum, a common thread that runs through these differences is the absence of F2F interactions that are familiar to F2F courses but unfamiliar to FO courses. As a result, some participants recognize a greater need to use their feedback to help students self-assess and remain engaged with the course. Additionally, where CAA feedback is expected to be used more in FO courses,
this feedback, at least as experienced with MML, is identified to be ‘more summative’ unless a student(s) initiates feedback interactivity. In summary, while participants contend the ‘ultimate’ purpose of feedback is the same in both contexts, findings suggest differences relative to the presence or absence of live F2F contact.

11.8.3 Quality of Learning and the Purpose of Feedback

As with findings concerning the use of discussion and quizzes, differences in the purpose of feedback continue to reflect a difficult environment where participants are attempting to adapt their feedback and assessment practices to meet these challenges.

Two participants’ accounts help elucidate this process of adaptation. P3 provides an interesting account of how s/he is adapting to the FO context which, when probed, s/he agrees that this FO course feedback is acting as surrogate for the absence of class meetings. Additionally, P6 shares how s/he is trying, albeit not very successfully, to adapt:

P3:  Frequently, what I do is I encourage them to hand in problems in pieces before the answers are posted. And they hand in problems and I look at them very carefully and give them lots and lots of feedback about what they’ve done wrong; with the hopes, that when they finally submit it for final grading, that feedback has helped their understanding and they can actually get it right.

P6:  ...sometimes I give feedback, but I don’t feel like it... generates the same sort of discussion so I’m kind of always wanting for that kind of feedback. And I don’t think my students read the feedback in the online course nearly as much... in a F2F course, the feedback is very verbal. You know I sit down with the student and say ‘You know, I looked at your homework and... You’re having a really hard time’...it’s more about the effectiveness of feedback.

In summary, participants identify several purposes for the use of feedback which, because of the absence of F2F contact, are realized differently in the FO as compared to the F2F context. In particular, compared to their F2F feedback, participants identify how their FO course feedback is being purposed to a greater degree for moving students to assess their own work and engage with the course material and each other. In addition, there is some suggestion that FO course feedback is ‘more summative’ whereas F2F course feedback is ‘more formative’.
11.9 Use of Feedback: Timing

**Study I Background:** Findings indicate that there appears to be no significant relationship between types of feedback used and any of the approach measures.

**Study II - R4:** How and why is feedback being provided? (Specifically, the timing of feedback used? In addition, do participants think that immediate feedback helps students develop their understanding of mathematics?)

Participants were asked to identify any differences in the timing of feedback in their F2F versus their FO courses. Additionally, they were asked whether they thought that immediate feedback helped students understand mathematics. Findings indicate that most participants both provide quicker feedback in their FO courses and view immediate feedback as providing some help but not for higher level understanding. This section first begins by detailing identified factors associated with feedback timing. Second, differences between the two contexts are covered. Third, participant-identified reasons are detailed as to why immediate feedback is and is not considered helpful for developing students’ understanding of mathematics. Lastly, findings as they relate to the quality of learning are summarized.

11.9.1 Identified Factors Associated with the Timing of Feedback

Participants identify similar factors affecting how they time their feedback as those already detailed with regards to the kind and process of feedback. In particular, the timing of feedback is seen to vary depending on whether CAA is used, the type of assessment instrument and the stage of the semester.

First, as suggested by the literature (e.g. Griffin & Gudlaugsdottir, 2006), the majority of participants’ responses associate the use of immediate feedback with the use of CAA and computer feedback. In particular, P4 and P5 refer to the use of CAA directly in response to the question on the effectiveness of immediate feedback:

*P3: My F2F course [homework using computer agency], the feedback is instantaneous.*

*P4: I also think a combination of the two, like some immediate feedback from the computer - yes, I think that’s good.*

*P5: Courseware - there is no question that that helps a certain kind of student.*
P6: ...my calculus and my college algebra courses have online [i.e. CAA] homeworks; so they get immediate feedback on their homeworks...

Second, two participants describe how their initial feedback timing varies according to the type of instrument – with responses showing differences in timing between exams, homework, quizzes and discussion.

P3: Well, in my calculus course I try to get my exams graded in two or three days and get them back to them... In my online courses it can take me up to a week to get the problems graded. [I: What about your online homework and your calculus?]... It’s instantaneous. [I: The last component was your quizzes?]... I try to give back to them the next class.

P6: ... they do quizzes in their [F2F] class and they do tests and that takes me a week, or a little over a week to get back. So with the discussions in my online course, I reply to discussions every day.

Third, one participant states how the timing of his/her initial feedback increases as the semester progresses:

P6: I do really well in the first couple of weeks of the semester - and then... the time just decreases, exponentially... So, I usually tell students on the first day of class that... it takes me about a week to get feedback back to them... And towards the end of the semester that will become two weeks - and that’s true of homeworks, exams – everything.

11.9.2 Differences in the Timing of Feedback

The majority of participants identify themselves as giving quicker feedback in their FO courses. While one viewed themselves as having ‘the same’ (P6) feedback timing in both contexts, the only participant providing quicker feedback in their F2F course (P3) is also the only participant whose FO course is dissimilar from their F2F course. Identified reasons given for differences in timing are two-fold: First, participants’ responses suggest FO courses are missing out on feedback that regular F2F class meetings provide both a space and time for. Second, students in FO courses are perceived to be more susceptible to frustration if they do not receive feedback promptly.

First, the major participant-identified reason relates to how FO courses lack the structure of regular class meetings, familiar to F2F courses, that provides a time and place for feedback interactions. That is, once again, the potential of F2F in-class feedback interactivity appears
highlighted – a potential noted by both two- and four-year participants – which is again consistent with typically smaller US class sizes. In addition, as previously discussed regarding differences in the use of instructional timing in discussion, some participants identify the need to provide feedback more quickly in the FO context. The response of P2 appears particularly insightful. In particular, where the F2F class is seen to ‘contain’ most or all of the feedback interactivity, the lack of a class in FO courses seems to open up those boundaries to an expectation of providing an almost ‘24/7’ interactivity. Finally, regarding pacing, s/he refers to regular F2F class meetings as a means of pacing students through the course whereas a quick response in FO courses is seen to fulfil a similar role:

P1: ...online; when I get an email I respond to it, assuming it’s during the hours I’m on campus, away or checking at home, as soon as I get it. When I’m dealing with students who are F2F, if they’re seeking personal feedback, the response is on the same time-frame, but since it’s most often feedback in the classroom, it’s dependent on the class schedule.

P2: I am online every day. I try to turn their questions around in 24 hours. A F2F class, most of my statistics courses meet twice a week, so we can go up to five days without seeing a student... in an online environment where students are engaging every day, [I think] that the instructor should do likewise... When a student in an online course has a question or an issue that they cannot self-resolve, they’re at a complete stand-still until they get assistance. And that is frustrating for an online student because there is no light at the end of the tunnel, there’s no ‘well Tuesday’s coming around the corner, I’ll ask then.’ So, I think that responding quickly to online students is part of pacing them through the course.

P5: Again, I think any sort of feedback from me is more important in the online class because I can give them informal feedback in a F2F class. If I have a class that meets four days a week and... [if the students have questions about the exam, I can go over the exam in that environment, we have more informal discussions. Whereas in the online class, that’s not really possible so I want to make sure the students get the formal feedback as quickly as possible...

P6: So with the discussions in my online course, I reply to discussions every day. At the end of the day, before I go to bed, every discussion post and every email has been read and replied to... [Later adding]... the difference is that... in the online course they watch the [instructional] video on their own time and they write a response to those discussion questions; whereas, in my [F2F] classroom, they watch the video and then we dialogue about it in the classroom.

Second, students taking FO courses are seen to become more easily frustrated with their learning if they don’t receive prompt feedback:
P2: And I also don’t want the students to get stuck. There can be a certain level of frustration in an online course if you don’t hear from your instructor when you have a problem... if there’s a real issue or concept of something that needs some attention, I just want the student to have attention very quickly.

In summary, in FO courses, participants’ responses identify the loss of regular class meetings with the lost potential for in-class feedback and/or interactivity. As a result of this loss participants seem to be providing feedback more quickly in their FO courses and, for some of these students, their learning experience is reported to be under stress.

11.9.3 Immediate Feedback and Mathematical Understanding

Participants’ responses differ as to whether immediate feedback helps in developing students’ understanding of mathematics. For example, two participants, the only two self-identified veteran instructors (P1 as ‘old school’ and P3 as ‘teaching for...thirty years’), consider immediate feedback as unconditionally helpful. Two others (P5 and P6), the community college participants, view immediate feedback, when provided within the context of repeated question attempts, as detrimental to the development of students’ understanding of mathematics. Various reasons are given as to why immediate feedback may or may not be viewed as helpful with participants, on balance, identifying its use as helpful for developing some kind of lower but not higher level understanding. Again, participants’ responses show a strong association, supported by the literature (e.g. Griffin & Gudlaugsdottir, 2006), between the use of immediate feedback and the use of CAA. This section begins by detailing those reasons why immediate feedback is considered helpful followed by those reasons why it is not.

Reasons Immediate Feedback is Considered to Help Students Understand Mathematics

Four reasons are identified for why immediate feedback is considered to help students understand mathematics: identifying errors, enabling learning processes, addressing affective needs and developing procedural understanding.

First, three participants identify the use of immediate feedback as beneficial due to the time proximity of error detection to the actual problem attempt:

P1: A word when you start to go off track, makes things much easier to fix then when you’re miles away from where you should be.

P3: It tells them right away whether they have done a problem wrong...
Second, two participants identify immediate feedback with the development of procedural skills:

\[ P2: \text{I think immediate feedback in terms of } \text{‘did you get this problem right or wrong?’ helps them with the mechanics.} \]

\[ P6: \text{...do the same procedure over and over and over again.} \]

Third, perhaps related to statistics as being a traditionally challenging course, two participants associate immediate feedback with what may be considered affective benefits:

\[ P1: \text{[Immediate feedback is considered beneficial because of]} \text{[p]ersonal experience of being so frustrated I’d like to throw the book across the room when I was in graduate school.} \]

\[ P5: \text{... motivation, for some students it enables them to engage with the material longer...} \]

Fourth, one participants’ response appears to associate the use of immediate feedback with enabling a cycle of learning:

\[ P3: \text{You know, if they’re doing a differentiation problem and they make a mistake. The problem will show them how to the differentiation step by step on that problem and then give them another differentiation problem. And that cycle repeats, at least for the good students, that cycle repeats until they can do the problem right.} \]

**Reasons Immediate Feedback is not Considered to Help Students Understand Mathematics**

Four reasons are identified for why immediate feedback is not considered to help students understand mathematics: works against higher- and reinforces lower-level understanding, limits the kind of questions to a focus on lower-level understanding and understanding is not retained and students are given the illusion of understanding without actually having it.

First, directly following P2’s previous statement regarding immediate feedback providing help with the ‘mechanics’, s/he qualifies that statement with the statement shown below. This statement reflects the theme most participants identify to be the dominant reason why immediate feedback does not help students understand mathematics. That is, that immediate feedback is seen to benefit the development of procedural skills to the detriment of developing conceptual (P2) or higher order (P5) understanding. Though limited by the predominant use of MML, participants’ responses are consistent with other findings discussed in the literature review.
concerning the use of CAA feedback (Morrison, Ross, Gopalakrishnan, & Casey, 1995), where some participants identify immediate feedback as working against or harming the development of deeper understanding, particularly within the context of iterative problem attempts:

P2: I don’t know that type of immediate feedback is going to help them... with a conceptual framework...

P5: I’ve seen students too often hitting on that help button to get them through a problem and I think sometimes that immediate feedback works against the student’s own understanding... my biggest problem... is that it reinforces low-level, procedural learning... I have a fully online course where I still ask higher-order questions but those do not have immediate feedback. It’s those... computer-adaptive systems where they get the immediate feedback, that’s where I think that procedural learning is emphasized and... conceptual understanding de-emphasized...

P6: I don’t think it [CAA with immediate feedback] helps them to understand what math is really about... they reinforce this idea of ‘Oh look! Here’s thirty ways to look at this problem.’ ... And if I’m stuck, I’m going to press this little ‘Help me solve this’ button and if I’m in a different homework system I’m going to press this ‘I can’t do this’ button. And it’s going to show me some professor who does exactly this procedure... THAT’S NOT WHAT MATH IS!!! ... math should have multiple solution paths and there should be some creative thought into how to get where I’m going... I really think online homework systems... reinforce a really bad concept.

Second, the use of immediate feedback is identified as limiting questioning to ‘lower level skills’ so that the questions may be quickly and easily graded (e.g. ‘open-ended’ questions are not seen to be viable)

P5: The questions they ask that... give immediate feedback are those questions that are easy for computers to score... they tend to ask very low-level skills...

For the final two reasons, P5 continues to elucidate ways s/he views immediate feedback as not helping students – singling out younger students in particular – develop their understanding of mathematics. The third reason extends the quality issue from one of depth alone to the longevity of any resultant understanding. S/he relates this to the previous issue regarding limited question types:

P5: [In reference to the use of immediate feedback with CAA] I have my doubts as to whether or not there’s a whole lot of retention, whether the students remember what they learn in these environments...when students... these take these assessments throughout the term... [and] then you have them get re-assessed maybe a month later... there’s virtually nothing there, it just sort
of evaporates. They learn these things in small packages, in small amounts of time and yet they don’t hold on to them...

For the fourth and final reason, P5 views immediate feedback as helping students develop what seems to be an illusion of understanding:

P5: [In summing up his/her view on the use of immediate feedback] ... in the long-run I think it’s a negative... You move away from your metacognition, thinking about your thinking, thinking about what it means and rather all your doing is hitting a button and see if it’s right or if it’s wrong... it gives them a false sense of security.

In summary, in some ways immediate feedback is seen as helping move students forward in developing their understanding of mathematics. However, in other ways it is seen as capping or even regressing any developing understanding. While the dynamic appears complex and, to some extent, are limited to MML, some participants (in particular P5 and P6) appear to single out the combined provision of immediate feedback with repeated attempts with a potentially serious and detrimental effect on the development of understanding.

11.9.4 Quality of Learning and the Timing of Feedback

As a whole, participants’ accounts of how they use feedback reflect a complex dynamic: one that, for example, is cyclical or interactive, involving the detection of erroneous answers followed by the timely provision of feedback. Within this dynamic, most participants provide quicker feedback in their FO as compared to their F2F courses. This seems to be largely a result of the lack of structure in FO courses leading some participants to try and provide almost ‘24/7’ interactivity.

In this respect, participants’ responses suggest that the quicker FO course feedback is in response to the loss of an instructional context where real time and F2F interactivity can be provided. And though quicker feedback may be of some benefit to student learning, it is the lost potential of real-time F2F interactivity that appears emphasized – as evidenced by at least one participants’ report of a student learning experience that is under increased risk of stress (i.e. frustration). However, recalling the issue of instructional timing covered in the section on the use of discussion, the challenge is not simply providing quicker feedback but judging and executing feedback timing for the purpose of realizing quality learning.
Finally, participants associate the use of immediate feedback with the use of CAA and hold differing views as to what, if any, benefit it has in helping develop students’ understanding of mathematics. While these views appear limited to MML, in terms of benefits, immediate feedback is seen to have largely surface benefits to the learning process (e.g. motivation) and some benefits to lower-level procedural learning, however no benefits to conceptual learning. Conversely, in terms of detriment, two participants identify the feedback interactivity provided by the combination of immediate feedback with repeated question attempts as detrimental to the quality of student learning. P5, for example, identifies this feedback as a ‘crutch’ that appears to be moving the cognitive processes away from the student to the computer. In a sense, moving students backwards when the purpose of feedback should be forward movement:

P5:  *I absolutely believe... that immediate feedback [within the context of repeated question attempts]... for many... becomes a crutch that they rely on that sort of mitigates against their real understanding...*

In summary, outside the use of CAA, despite some potential benefits of quicker feedback provided in FO courses, the balance of evidence suggests that without real-time F2F interactions it is difficult in FO courses for participants to optimally time feedback provision in order to effectively realize quality learning. Moreover, with the use of CAA and immediate computer feedback and limited by participants’ predominant use of MML, these systems are viewed to, at best, lead to lower-level learning and, when coupled with repeat attempts, be possibly harmful to any developing understanding.
Table 33: Summary of second study results

<table>
<thead>
<tr>
<th>Assessment Practice</th>
<th>Factors Affecting Use</th>
<th>Emergent Themes</th>
<th>Differences Between F2F and FO Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
<td>1. Nature of Communication – F2F discussion as natural and FO discussion as unnatural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Monitor the state of student thinking – as a means of informing the instructor if feedback is needed to correct student thinking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Build and maintain community</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Help deepen student understanding</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>1. Nature of Communication</strong> – F2F discussion as natural and FO discussion as unnatural</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>2. How Discussion is Formalized</strong> – FO appears much more likely to be attached to specific task and grade than discussion in F2F courses</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>3. Amount of Communication Resources</strong> – Less in FO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>4. Use of Instructional Timing</strong> – more difficult in FO to both judge and execute the appropriate instructional timing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>5. Use of Interpersonal Dialogue</strong> – Loss in FO of an iterative or cyclical instructional dialogue (e.g. student-instructor-student...) that occurs in the moment of, and considered necessary for, effective instruction.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>6. Use of Collaboration</strong> – Less in FO</td>
<td></td>
</tr>
<tr>
<td>Quizzes</td>
<td>1. Nature of the F2F and FO context (e.g. ease of using invigilation in FO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Nature of mathematics (i.e. level of complexity)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Availability of resources (e.g. CAA packages that accompany texts)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1. Monitor the state of student thinking</strong> – as a means of informing the instructor if the instruction needs to be altered</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2. means of directing students through the course material in at least three ways: engagement, pacing and transitioning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1. Use and Value - FO for monitoring and pacing student activity. F2F for monitoring and challenging student understanding.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2. Flexibility - less in FO (e.g. changing questions and how they are administered)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>3. Invigilation - expectation FO quizzes not invigilated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>4. Feedback provision – greater likelihood via recording in FO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invigilation</td>
<td>1. Threat posed by the widespread use of cheating (e.g. ‘rampant’ as described by P1) where mathematics courses are reported to be particularly at risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Internal and external interests for and against (e.g. mathematics department in favour while the administration is against)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1. Validation of student learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2. Learning benefits when, for example, half the course exams are invigilated and the other half are not</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1. Required Resources – much more in FO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2. Kind of Invigilation – more formal in FO. F2F may involve a ‘degree’ of invigilation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback Kind</td>
<td>1. Student demographic characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. The type of assessment question (e.g. solving a problem versus providing a definition of a mathematical term) or instrument used (e.g. exam versus homework)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. The context of where assessment feedback is provided. For example, whether submitted assessment instruments receive feedback directly on the assessment instrument (e.g. written feedback), in-class (e.g. verbal) and/or on an</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1. Amount of Feedback Provided – More in FO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>2. Provision of Feedback: By Question or Instrument – With use of CAA, more by question in FO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>3. Provision of Feedback to the Entire Class: In-Class F2F vs. Class-Wide FO – Initial feedback may be better in FO but feedback interactivity less effective</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>4. Provision of Feedback to</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>Timing</td>
<td>Identified Most Effective Kind of Feedback:</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1. Use of CAA (usually provides immediate feedback) 2. Type of instrument (e.g. longer for exams than quizzes) 3. Stage of the semester (e.g. slower as the semester progresses)</td>
<td>1. Forward movement 2. Develop ‘understanding’ 3. Identify errors 4. Helping students self-assess 5. Stimulate and maintain interactions</td>
<td>1. Process: ‘one-on-one’ or ‘side-by-side’ 2. Process: Interactive 3. Process: Immediate</td>
<td></td>
</tr>
<tr>
<td>1. Identify errors 2. Summative Assessment Emphasis – More in FO because of the greater dependence on CAA</td>
<td>1. Structure of Instruction – Quicker in FO because F2F class is seen to ‘contain’ most or all of the assessment feedback interactivity while the lack of a class in FO courses seems to open up those boundaries to an expectation of providing an almost ‘24/7’ interactivity 2. Students’ Affective Needs – Quicker in FO where students are more easily frustrated if they don’t receive prompt feedback</td>
<td>4. Kind: ‘correct’ thinking (e.g. right/wrong, full solution) as being the least effective and feedback that provides hints, comments or questions student thinking</td>
<td></td>
</tr>
</tbody>
</table>

**Identified Reasons Immediate Feedback is Considered to Help Students Understand Mathematics:**

1. Time proximity of error detection to the actual problem attempt
2. Development of procedural skills
3. Affective benefits (i.e. reduces frustration, increases motivation)
4. Enables a ‘cycle’ of learning

**Identified Reasons Immediate Feedback is not Considered to Help Students Understand Mathematics:**

1. Works against higher- and reinforces lower-level understanding
2. Limits the kind of questions to a focus on lower-level understanding
3. Understanding is not retained
4. Helps students develop what seems to be an illusion of understanding or ‘false sense of security’ that they understand the mathematics
12. Study II Discussion

Overall findings illustrate the struggles these six US public HE instructors are experiencing in adapting their assessment practice to the FO course context. They suggest that, because of the loss of F2F synchronous interactions, they are struggling to adapt their assessment practice, particularly in a way that leads to deeper understanding. Findings highlight differences between the efficacy and flexibility of F2F pedagogy and the more rigid and formal FO pedagogy.

This chapter discusses the findings for each of the five research questions posed in the second study. The discussion for each of these questions is structured as follows. First, general findings are summarized for both the first and second thesis studies and related to the literature. Second, any limitations are summarized. Third, the discussion concludes with a summary of answers to the research questions. Following this, in the concluding chapter, some of the findings are related to current theoretical arguments in distance education.

12.1 How and why is discussion/interaction used?

The findings from the second study indicate three purposes for which discussion is used—monitoring the state of student thinking, building and maintaining community and helping to deepen student understanding. These are consistent with prior general findings (Parisio, 2010). The first study indicated that fewer than half (39%) of the participants used discussion as a weighted component of assessment (mean weighting of 10%) in their FO courses. This weighting is consistent with prior findings, however the percentage of instructors using discussion is considerably lower (compared to 78 percent using discussion weighted at 7 percent, Galante, 2002; and 70 percent using discussion weighted at 13 percent, Trenholm, 2007a). Similarly reflecting a decline, SUNY participants’ use and weighting of discussion in this thesis was found to be significantly less than SUNY participants’ use and weighting of discussion in the 2007 study (Trenholm, 2007a). This is also consistent with a trend away from ‘community orientation’, reported from 2002 to 2007, in FO courses in the pure disciplines at one major SUNY university (Smith, Heindel et al., 2008).

In this second study, participants’ accounts of using discussion appear consistent with expectations, because of typically smaller classes than what may be found in some other countries, that they are accustomed to having opportunities for regular class discussion. In particular, due to typically even smaller classes and an emphasis on teaching, the two
community college participants’ accounts appear consistent with instructors who may be more sensitive to the impact of the affordances and constraints of the FO environment on their teaching abilities. However, as a weighted component, only three of the six participants used discussion in their FO course while only one participant reported doing so in their F2F course. And for both weighted and non-weighted discussion, while overall findings indicate that participants identify some benefits to discussion in FO courses (e.g. more time to articulate thoughts), all experience more difficulty in using discussion in their FO than their F2F courses.

Moreover, the analysis of differences in the use of discussion in F2F and FO courses indicates that participants are struggling with six issues. Four identified differences – FO course discussion identified as unnatural and formal and the FO course context seen to present a difficult environment to effectively use instructional timing and practice collaboration – appear to be largely symptomatic of two other differences: the loss of a real-time iterative or cyclical instructional dialogue and poor communication resources. This section focuses particularly on these two root issues and relates these findings back to the literature. Finally, the second study provides one plausible explanation for the lack of association found between the study measures and the use of discussion in the first study. That is, despite any inclination to engage students in knowledge construction, at least one of the second study participants opted out of using discussion in their FO course due to the challenges discussed.

First Root Issue

The first of the two root issues is the identified cyclical instructional dialogue (e.g. student-instructor-student...) that occurs in the moment of, and is considered necessary for, effective instruction (e.g. P6: ‘it’s lacking that synchronous give and take’). This kind of process in mathematics instruction is consistent with claims about how mathematics is learned by alternating discussion with reflection (Elbers and Streefland, 2000; Elbers, 2003). Indeed, the need for periods of reflection has been linked to developing students’ understanding (Janvier,1996; Goodell, 2000). As Skemp (1979) conjectured, higher-level thinking appears to be encouraged by an ‘alternation’ between reflection and discussion. In summary, consistent with the literature, participants’ responses suggest a kind of reflective interactivity is needed to effectively help develop students’ understanding of mathematics.

However, participants’ responses suggest the reflective interactivity needed for mathematics instruction is not being experienced in their FO courses because the asynchronous nature of FO courses does not easily permit it (e.g. P2 and P6 refer to F2F but not FO discussion as ‘natural’).
This would appear to be supported by prior research at the school level which places the cycle of ‘alternation’ (Skemp, 1979) for quality discussion at no more than 3 to 5 seconds (Tobin, 1986). Moreover, it is in stark contrast to the FO course context where in general, for example, only 10% of instructors view the ‘ideal frequency of required discussion postings’ to be more than twice a day37 (Mandernach, Gonzales, & Garrett, 2006, p.254). These differences are illustrated in Figure 14.

Furthermore, this argument finds support in the literature where researchers identify a missing level of interactivity in distance education and FO mathematics courses. For example, Kloeden and McDonald’s (1981) early survey research found that the most common difficulty expressed by distance learning mathematics students was ‘insufficient detail’, which they argued may be partly due to the absence of a lecturer who can ‘‘adlib’ the missing steps in the lecture’ (p.60). More recently, in comparing F2F and FO mathematics courses taught in the Asian context, Ramasamy (2009) notes missed opportunities to receive step-by-step instruction as well as to ‘inject’ questions when doubts arise, concluding these differences are serious enough as to ‘jeopardize’ the development of students’ understanding (p.5). Finally, in the UK context, Foster (2003) conducted a survey and found that while his participants see FO course interaction as complementing F2F interactions, ‘there is no evidence they can replace it’ (p.148). In summary, the experience of interview participants is consistent with the literature and contrasts the more halting flow of reflective interactivity available in FO courses with the more continuous flow identified as needed for mathematics instruction.

37 These statistics are for an 8-week accelerated course. 67% participate viewed the ideal frequency to be daily to one day per week and 23% did not believe there should be any requirements for participation.
One final characteristic of reflective interactivity in mathematics relates to the role of the instructor. The kind of interactions participants identify as needed suggests that FO discussion would be more expert- rather than student-led\(^{38}\). This is consistent with claims about mathematics instruction in general (Wong, 2009) and e-learning in different disciplines (White & Liccardi, 2007), as well as the problem of fit, discussed in the literature review, regarding the emphasis on student-led discussion in FO instruction being incompatible with mathematics instruction.

In summary, participants’ responses highlight the nature of reflective interactivity needed and the considerable challenge being experienced in attempting to effectively develop students’ mathematical understanding in the FO course context. Where some have claimed the FO course context enables students to engage in reflection that leads to quality learning (general FO literature e.g. Swan, 2001; mathematics education literature e.g. Rosa & Lerman, 2011), these findings appear to call these claims into question.

---

\(^{38}\) This does not preclude the use of ‘student-experts’. Though the assumption is that even in these instances, as indicated by participants’ responses, the instructor remains an active observer ready to correct any ‘incorrect thinking’ (P2).
Second Root Issue

The second of the two root issues identified by participants, poor communication resources, appears to be compounding the first root issue. Findings support claims in the literature concerning the need for multiple modes of communication in mathematics, such as verbal cues, gestures, facial expressions and silence. In particular, the lack of communication resources in discussions is making it more difficult for participants to judge and monitor the state of student understanding and provide appropriate instruction (Skemp, 1976; Pirie & Schwarzenberger, 1988). As P6 states: ‘I really feel like there’s a lot more to communication than just text and just words... all the other parts of communication [are missing]...the gesture...the graphic’. While such a challenge has been noted in other disciplines (Hrastinski & Stenbom, 2012), some claim a greater need exists in abstract disciplines such as mathematics for, for example, gestures and diagrams which act as a cognitive support to learning (de Freitas & Bentley, 2012; Edwards, 2009; Roth, 2001). As Lowrie & Jorgensen (2012) recently found, distance mathematics educators prefer real-time verbally-based (vs. text-based alone) two-way communication. This issue will be further discussed when use of feedback is covered.

Finally, in relation to both of these root issues, it seems understandable that participants may identify the use of FO discussion as, for example, unnatural, more formal and incompatible with collaborative mathematics activities. Issues such as these suggest reasons why instructors, despite any inclination to engage students in knowledge construction are not requiring discussion in the form of a weighted assessment instrument in their FO courses.

12.1.1 Limitations

There are at least a few limitations related to answering the current research question. First, as shown in Table 28, while three participants include the use of discussion in their FO assessment scheme, only one includes discussion in their F2F assessment scheme. It is unclear what effect allocating a portion of the course weighting to the use of discussion has on how instructors and students engage in discussion. Does this, for example, lend to the experience of unnaturalness identified by participants if students feel forced to participate? Second, findings suggest students have different choices regarding their participation in F2F compared to FO discussion that limit how instructors can orchestrate discussion. For example, when attending a F2F class students choose whether or not to ‘tune out’ (P1) of any discussion. However, in the FO course context, students must first choose to enter a discussion before the option to tune out arises.
That is, in the F2F course context, there is a much greater likelihood that students will be either a silent witness or an active participant in any in-class discussion. In contrast, in the FO course context, students may be ‘lurking’ yet it is impossible to know what they are thinking. That is, the choices available to students suggest it would be more challenging for participants to engage students in discussions in the FO course context. Third, the nature of discussion remains unclear (i.e. how much is ‘purposeful talk’, see Pirie & Schwarzenberger, 1988, p.461) – for example, how much of discussion is addressing students’ behavioural issues versus their thinking in mathematics. Here again, F2F discussion is favoured, for example, when instructors are able to simultaneously walk around, monitor and direct an entire class discussion. In contrast, to be able to monitor and direct an entire FO course discussion, all student FO discussion contributions must be read individually. Such limitations highlight some of the complexities and possible confounds involved with comparing F2F and FO discussion.

12.1.2 Summary

Evidence of participants struggling or even opting out of using discussion confirms the problem of fit regarding the use of discussion. Though, as discussed in the literature review, a problem with the mechanics of communicating was expected to be a dominant issue of fit, the second study suggests two other primary issues. They suggest that current FO discussion is not providing these participants with the context for the kind of reflective interactivity that is needed to effectively help develop students’ understanding in mathematics, and is compounded by a lack of communicative resources such as the use of gestures or silence. Where the importance of discussion has been highlighted in both FO (Swan, 2001) and F2F (Skemp, 1979; Wood, 1988; Shepard, 2000) teaching and learning, this suggests a degraded experience in FO mathematics courses and a possible reason why the rate of attrition is so high in FO mathematics courses (e.g. Smith & Ferguson, 2005).

12.2 How and why are quizzes being used?

Participants’ responses suggest two purposes for using quizzes – to monitor the state of student thinking and as a means of directing student activity – that are consistent with prior findings (Griffin & Gudlaugsdottir, 2006; Varsavsky, 2004). The first study found that a little more than half (53%) the participants used quizzes as a weighted component of assessment (mean weighting of 20%) in their FO mathematics courses with none of these quizzes invigilated. These findings contrast with expectations from the literature where the use of quizzes is
considered a central characteristic of FO instruction (Engelbrecht & Harding, 2005a; Greenberg & Williams, 2008).

In this second study, as a weighted component, only three of six participants used quizzes in their FO course (two CAA- and one paper-based) while all participants reported using them in their F2F course (one CAA- and five paper-based). All of the F2F but none of the FO quizzes were invigilated. Similar to the use of discussion, overall findings indicate that participants are having more difficulty using quizzes in their FO than their F2F courses. The analysis of differences in F2F versus FO use of quizzes indicates that participants are struggling with a few issues that appear inter-related. In particular, the lack of invigilation, reduced pedagogical flexibility and the nature of feedback in FO courses suggest possible reasons why FO quizzes are identified more for their use in directing than advancing student learning. Leaving the use of invigilation and feedback for later in the second study discussion, this section begins by relating the two remaining differences (i.e. pedagogical flexibility and the emphasis on directing and advancing student learning) back to the literature. Then, given a dependence on CAA in FO courses and with the caveat concerning the predominant use of MML, most of the discussion focuses on the identified nature of learning resultant from CAA-based quizzes (termed ‘online quizzes’).

First, participants identified a lack of flexibility in the way quizzes are administered in FO courses as compared to F2F courses that suggests a reduced ability for using quizzes as a tool for advancing student learning. In particular, participants identify having the choice in their F2F courses, as they judge to be needed, to administer quizzes either to be completed individually or collaboratively in small groups. In contrast, in their FO courses, the expectation appears to be that quizzes should only be completed individually. Such a difference suggests FO courses are missing out on a valuable tool for developing higher quality learning (Goos, 2004). Moreover, in response to perceived student needs (offering either more challenging questions or questions that help transition students to new material), participants identify an ability in their F2F courses to change questions, for example, just prior to administering a quiz. In contrast, consistent with prior findings, FO courses lack the ability to adjust their pedagogy in an ‘impromptu fashion’ (Harman & Dorman, 1998, p.307). In summary, such a loss of pedagogical flexibility suggests the use of quizzes has less potential for advancing student learning in FO than F2F courses.

Second, the use of quizzes is also identified as more of a tool for directing rather than advancing student learning. This is consistent with claims that online quizzes are being used to ‘encourage
(or even force)” (Griffin & Gudlaugsdottir, 2006, p.486) students to ‘work consistently throughout the semester’ (Varsavsky, 2004, p.167). However, it is unclear whether, for example, this emphasis on directing learning in FO courses is due to concerns regarding the validity of students’ quiz grades (due to them not being invigilated).

**Issues Related to FO Course Dependence on CAA**

However, perhaps the most significant issue regarding the use of quizzes relates to participants’ reliance on CAA. In the first study, of all participants using quizzes in their FO courses, just over four fifths (26/32 or 81%) report using online quizzes and, of these, almost all provide immediate feedback (24/26 or 92%). Yet, limiting any potential discussion on the use of CAA in general, most participants appear to be using the same CAA system; MyMathLab (MML). With this limitation in mind, two themes emerge from the second study results which together suggest online quizzes, at best, help students gain lower-level procedural knowledge.

First, online quiz questions are identified as only being capable of addressing and facilitating lower-level learning. For example, P3 refers to students only being asked to do ‘trivial’ questions or ‘computations’. Similarly, P5 refers to an inability to ask ‘open-ended’ questions which have been linked with deeper learning (Scouller, 1998). This suggests a possible explanation for the positive correlation between quiz weighting and an information transmission/teacher-focused approach (ITTF) – as measured by the ATI. That is, given quizzes are associated with lower-level learning, it may be expected that those with an approach more associated with this kind of learning (i.e. surface approach; Trigwell & Prosser, 2004) will place a greater emphasis on the use of quizzes and conversely, those with an approach less associated with lower-level learning may be expected to place less of an emphasis on quizzes.

Second, suggesting a detriment to learning, some participants provide a sense that when the scaffolding provided by CAA systems is removed the constructed knowledge does not remain standing. For example, P5 refers to knowledge that is learned in ‘small packages, in small amounts of time’, not ‘necessarily internaliz[ed]’ which ‘sort of evaporates’ over time. Similarly, P4 states that s/he wants to ensure his/her students can ‘do it on their own’ without the help offered by CAA (e.g. MML’s ‘Help Me Solve This’ resource). Reflecting on current theory, these characterizations are consistent with procedural knowledge that is not deep but

---

39 Participants were considered to be using CAA if either they reported using immediate feedback and/or computer-generated feedback. The remaining participants (6/32) are assumed to have electronically administered (e.g. a downloadable document) their own paper-based quizzes.
‘connected exclusively or largely with other nonconceptual knowledge... [and that] tends to yield more error-prone, rigid, short-term or isolated extensions’ (Baroody et al., 2007, p.126). Additionally, reference to a focus on ‘small packages’ (P5) or ‘little bits of topics’ (P1) may suggest, according to the literature (Baroody et al., 2007), that these quizzes are not helping students deal with the ‘big ideas’ that help connect concepts they are learning. Overall, some participants contend that some students become dependent on CAA to be able to demonstrate that they understand the mathematics, giving some students a false perception that they understand the mathematics when they actually do not.

In summary, the findings appear mostly limited to the use of MML. They support claims that question the actual value of online quizzes (Paterson, 2002) and contradict claims that they help ‘consolidate’ understanding (Lowe & Hasson, 2011, p.40). Furthermore, participants’ responses are consistent with recent findings that suggest MML may have ‘a negative effect’ on some students (Radu & Seifert, 2011) and contrast with MML in-house research literature that claims its use results in ‘enhanced instructor effectiveness... and improved learning’ (Speckler, 2010, p.64). The use of CAA will be further discussed in the section on the use of feedback.

12.2.1 Limitations

There is one significant caveat regarding the analysis of participants’ stated experiences of using CAA. As previously discussed, four participants use or refer to MyMathLab (MML) when answering the interview questions (specifically referred to by P1, P2 and P4 and indirectly by P6). This proportion of MML users is not be surprising given the share of the market MML currently holds with regards to US HE mathematics CAA software (which was 46% and growing in 2006; Monument Information Resources as cited in Freestone, 2006). However, the predominant use of MML by participants’ suggests their views on the use of CAA are mostly limited to MML. In addition, apart from some commonly associated functionalities discussed in the literature (e.g. immediate feedback and multiple attempts with CAA-based quizzes, in general, and, as will be discussed, MML specifically), current CAA systems are assumed to be complex with the nature of feedback and functionality offered varying considerably between systems. Moreover, with ongoing advances in both software and hardware these capabilities are considered to be constantly changing. In this sense, findings may be seen to reflect more on MML than on the use of CAA.
12.2.2 Summary

Findings are limited by the interview participants’ dominant use of a single CAA system, MML. They suggest that, when adapting to the FO context, interview participants are using quizzes less. As argued earlier in this discussion, this appears related to reduced flexibility in how quizzes may be used in FO courses and that these quizzes are not invigilated. In relation to student learning, in the FO context it is unclear from participants’ responses whether the use of quizzes, which two out of the three interview participants report to be CAA-based, has any benefit beyond monitoring and directing student learning activity. However, in contrast, in the F2F context, where almost all interview participants report using paper-based quizzes, responses suggest ways these quizzes may be used to advance student learning by, for example, being changed in situ in response to perceived instructional needs or used collaboratively in a way that resembles the use of constructivist pedagogy (Richardson, 2003). On balance, whereas F2F course quizzes have the potential to be used flexibly for advancing student learning, it is unclear what benefit FO course quizzes have beyond directing student learning.

12.3 How and why are participants choosing to use invigilation?

The second study found one primary purpose for which participants used invigilation – the validation of student learning – that is consistent with prior claims (Trenholm, 2007b, Flesch & Ostler, 2011). The first study found that about two-thirds (62%) of participants use some form of invigilation in their FO mathematics courses, contrasting with previous findings where about one-third (36%) used some form of invigilation (Trenholm, 2007a). Additionally, the first study participants who did not use invigilation were found to use significantly more assessment weighting associated with instruments that provided hints and comments (type 2 feedback) and have a significantly lower S&B measure. This raised some question about the nature of feedback provided by non-invigilating participants given better quality feedback has been linked to KC (Butler & Winne, 1995).

In this second study, four of six participants used invigilation in their FO course while all used invigilation in their F2F courses. Of those who did not use invigilation in their FO course, one did so claiming accessibility issues for disabled learners and challenges of having ‘students around the world who we can’t physically get together’ (P1). The other participant did not invigilate because of the nature of their course – which taught students how to use Mathematica – and claimed they had the ability to detect ‘by the form of their answers whether they've
worked together [cheated]’. Differentiated from the use of discussion and quizzes where the focus is on the added value to student learning, the primary focus of the use of invigilation is whether the learning has actually taken place – that is to what extent this grade reflects a student’s capacity to think mathematically or their dependence upon animate (e.g. another person) or inanimate (e.g. text) resources (Trenholm, 2007b).

In considering whether to use invigilation participants identify two principal factors that are consistent with prior claims. First, there is a general threat posed by the widespread use of cheating (McCabe, 2005) where mathematics courses are seen to be particularly at risk (Trenholm, 2007b). For example, P4 reported the ease with which students ‘cut and paste’ discussion portions in order to satisfy assessment requirements. Second, influences both internal and external to mathematics departments seek to either eliminate or ensure the use of invigilation (Flesch & Ostler, 2011). For example, offering reasons why some FO mathematics instructors are not using invigilation, P6 reports how their mathematics department is resisting pressure from the administration while P4 reports how their mathematics department is resisting pressure from departments outside of mathematics. These accounts suggest a conflict between those most intimately aware of how mathematics is taught and learned and others inside and outside mathematics departments. In particular, relating to P6 as a community college educator, his/her account is consistent with a conflict between institutional goals of accessibility and the hurdles introduced by requiring invigilation.

The analysis of differences in F2F versus FO use of invigilation indicates that participants are struggling with three issues. First, as others have found (Prince, Fulton, & Garsombke, 2009), the use of invigilation in FO courses requires more time, and human and physical resources. Furthermore, participants’ responses suggest differences in required resources between local and remote FO students. Second, participants’ responses suggest an expectation that all assessment instruments administered in the F2F class context offer some level of invigilation. This is contrasted with FO courses where no invigilation may be used but when used appears to be a much more formal process. Third, as others have claimed (Trenholm, 2007b), parties internal and external to the mathematics department are interested in whether FO mathematics courses are completely non-invigilated. Perhaps because there is no impact on accessibility, the same interest is not identified for F2F courses.

Finally, as compared to those survey participants who reported using invigilation, findings suggest a possible reason why those not using invigilation use significantly more assessment
weighting associated with instruments that provided hints and comments (type 2 feedback), despite having a significantly lower S&B measure. That is, looking to results concerning the differences in the purpose of feedback, this may be feedback that fulfils an intermediate purpose such as helping students self-assess and remain engaged with the course. Alternatively, as a question for future research, it may be part of an effort at addressing concerns about student validity.

12.3.1 Summary

Results regarding the use of invigilation introduce an additional level of complexity to researching current FO mathematics assessment practices (Iannone & Simpson, 2011). Where traditional F2F invigilation practice appears largely assumed and unchallenged, findings suggest it is relatively easy for FO mathematics instructors to choose not to use invigilation given, for example, pressure from administrators consistent with providing accessible education and combined with the potentially onerous time requirements for administering invigilated exams to remote students. While some instructors choose not to use any invigilation, the identified level of resistance by other participants and their mathematics departments suggests the stakes are significant and the issue of whether invigilation should or should not be required is unresolved.

12.4 How and why is feedback being provided?

The findings from the second study indicate five purposes for which feedback is used – forward movement, developing student ‘understanding’, identifying errors, helping students self-assess, and stimulating and maintaining interactions. ‘Forward movement’, in particular, was identified by every participant. This purpose, together with the second, is consistent with probably the most well-known purpose of feedback cited in the literature – to facilitate closing the gap between a student’s present and a desired state of learning (e.g. Hattie & Timperley, 2007). The purpose of identifying errors is consistent with an emphasis on ‘correct answers’ (Havnes et al., 2012; Hodgen & Marshall, 2005) while the latter two purposes are consistent with claims regarding the purpose of feedback in FO courses (Comeaux, 2005 as cited in Austin, 2007; Gikandi et al., 2011). The purpose of helping students self-assess appears to extend current findings in the literature and is consistent with the identified loss of communication resources linked with the use of discussion. That is, where F2F instructors may typically access and respond to student thinking by ‘reading’ students’ faces in a class setting, they now require students to actively recognize and communicate their own state of thinking.
The first study found that, on balance, participants emphasize the use of type 2 feedback with most (48/66 or 73% of all participants) using a combination of two or more types of feedback with their assessment instruments. Findings suggested problems with feedback literacy given that no significant relationship was found between the three feedback and any of the three approach measures. This appeared to be supported by results of Cronbach’s alpha test – run on the S&B questions – that indicated participants’ approaches to using feedback (i.e. the last S&B question) was least related to their overall approach to assessment. Questions of feedback literacy continue to arise in the second study when participants almost consistently refer to feedback processes even though specifically asked about kind(s) of feedback – suggesting participants connect the quality of their feedback practice primarily with how and not, as emphasized in the literature (e.g. Hattie & Timperley, 2007), with what kind of feedback is provided.

To answer the research question how and why is feedback provided, three separate interview questions were asked to cover three aspects of participants’ use feedback – the kind (process), purpose and timing of feedback. Illustrating the complex nature of feedback provision, the analysis of differences in F2F versus FO use of feedback for all three questions produced 11 participant-identified differences – six for kind, three for purpose and two for timing. These differences suggest participants see two stages related to the use of feedback: a first stage where initial feedback is provided to an individual student after completion of an assessment question or instrument and a potential second stage where generally a follow up assessment discussion uses the initial feedback as part of an instructional dialogue most often associated with class-wide discussion. This distinction is consistent with recent research that has focused on assessment ‘conversations’ (Ruiz-Primo, 2011), ‘discourses’ (Björklund Boistrup, 2010) or ‘discussions’ (Dunn & Mulvenon, 2009). Overall findings suggest that while initial instructor-provided feedback may be qualitatively better in FO than F2F courses, follow-up assessment discussion is qualitatively better in F2F than FO courses. As with the use of discussion, these findings suggest participants are having problems executing the necessary reflective interactivity in FO courses, including making use of multiple modes of communication. Finally, for both feedback stages, limited by the dominant use of MML, findings suggest instructor-provided feedback is qualitatively better than CAA feedback.

---

40 Other examples include, for example, a student(s) asking for clarification about a question on an assessment instrument.
Along each of these two stages – initial feedback and follow up assessment discussions – the following sections consider F2F compared to FO instructor-provided feedback and then CAA-provided feedback.

**A Comparison of F2F and FO Instructor-Provided and CAA-Provided Feedback**

Comparing F2F and FO instructor-provided feedback, for the first stage, participants’ use of initial feedback by may be summarized as follows: In their FO as compared to their F2F courses, participants are providing more individualized, detailed and quicker feedback. However, consistent with prior claims (e.g. Simpson, 2002 as cited in Lingefjärd & Holmquist, 2002), participants state that extra time and effort is required to give feedback in their FO as compared to their counterpart F2F courses. In the second stage participants’ responses suggest, consistent with previously discussed expectations related to the size of US HE mathematics classes as well as the community college focus on teaching, that they place considerable value on discussing assessment feedback. These responses provide further evidence suggesting the FO course context is not addressing the nature of reflective interactivity needed for mathematics instruction. First, recalling participants’ responses on the use of discussion, the ‘practice of mathematics requires frequent feedback’ (Kantor, 2003, p.1) and differences in interpersonal dialogue suggest the feedback loop (Hounsell et al., 2008) in mathematics instruction is tight – a ‘tightness’ that participants seem to experience as much more available in their F2F than their FO courses. Second, what participants’ regard as the ‘most effective kind of feedback’ (i.e. ‘side-by-side’, interactive and immediate) and how they link immediate feedback to developing students understanding (i.e. time proximity of error to feedback and facilitating a cycle of learning), as well as how they are providing their FO course feedback more quickly than their F2F feedback, all suggest a kind of interactivity that participants may more readily achieve in their F2F than their FO courses. Together these findings suggest that while initial feedback may be better, though requiring more time and effort in FO as compared to F2F courses, assessment discussions are not working as well in FO as compared to F2F courses.

What seems to be an obvious reason why these discussions are not working as well in FO courses is that participants are having difficulty leveraging initial FO course feedback for instructional purposes in subsequent interactions with students. As the use of initial assessment feedback for instructional purposes suggests, assessment, instructional and learning processes are all intertwined (Ramsden, 2003; Smith, 2002). Moreover, as both P4 and P6 describe the F2F course context as having the potential to be ‘formative’ whereas, in FO courses, the use of
feedback is identified as ‘more summative’, this appears consistent with problems moving from stage one to stage two feedback. In summary, participants’ accounts appear to highlight the struggles with mediating for the loss of a feedback process where initial feedback may be used as an opportunity to extend learning.

Moreover, these challenges may be further understood by considering the initial feedback and assessment discussion stages together in terms of three feedback steps. The first step is considered to be judging evidence of student learning, the second step is considered to be selecting appropriate feedback and the third step is considered to be the timing of feedback provision.

In the first feedback step, the judgement of student learning in FO courses is challenged when evidence is limited to text rather than the multiple sources common in the F2F course context – what Edwards (2009) refers to as an ‘enhanced’ ability for instructors to judge students’ thinking (p.139). In the limited FO literature, findings are consistent with what Harman and Dorman (1998) state with regards to distance teaching of undergraduate mathematics: compared to F2F instruction ‘it was more difficult to pick up visual and eye-to-eye cues and responses’ (p.307). As Tall (1977) also claimed, in the general mathematics literature: ‘annoyance, fear, or just a dull lost look in the eyes... are all signs of the state of the brain’ that should not be separated from ‘the cognitive side of learning’ (p.11; italics mine). Yet in FO courses they are separated when the instructor cannot see his/her student(s). P6 describes the situation this way: ‘[In my F2F class] I could see their faces and they can see mine... [In my FO courses] I build in some extra assessments to make sure that they really are getting what they're getting cause I don't have that intuitive assessment that happens when I'm sitting next to them’. Participant experiences are consistent with current research that suggests multimodal sources for judging students’ thinking is particularly important for mathematics instruction (e.g. Roth, 2001). Even evidence as mundane as silence is to be considered a window into the state of student thinking (Björklund Boistrup, 2010). In effect, the literature suggests, as Tall’s claim would seem to imply, that FO course instructors and students are disadvantaged when, for example, the only available evidence of learning is text-based.

In the second feedback step, this evidence is then used to make ‘crucial’ inferences that help instructors select appropriate feedback where, if those inferences are wrong, some claim learning is ‘less likely to occur’ (Bennett, 2011, p.14). Findings suggest two issues related to the selection of appropriate feedback. First, there appears to be a greater possibility that instructors
make an inaccurate judgement of student learning which leads to a wrong inference and thus a greater likelihood that less appropriate feedback is provided. Second, a prior state of feedback illiteracy, as suggested from both studies, may be exacerbating these issues. In summary, given the trickle-down effect of poorer evidence, discussed in the first feedback step, the potential of selecting inappropriate feedback appears to be greater in FO than F2F courses.

In the third feedback step of timing feedback, participants’ responses suggest that it is difficult in FO courses to time feedback provision to be most helpful to student learning (e.g. P3: ‘There’s a tremendous difference; in the sense that when I’m teaching [F2F]... I pause – I ask a question – and then I pause...’). Given the literature suggests there is an optimal time to provide feedback (Brookhart, 2004; Gibbs and Simpson, 2004; Havnes et al., 2012) such that, for example, the greater the need for students to process the material to gain understanding, the more benefit they derive from delayed feedback (Hattie & Timperley, 2007), this is relevant to what this thesis argues concerning reflective interactivity. For example, the frequency of interactivity in FO courses is likely to mean that by the time a student has reflected on an instructor’s comment that challenges his/her understanding, s/he is likely to have moved on to another concept. What effect this has on how students learn the conceptual structure of mathematics (Richland, Stigler, & Holyoak, 2012) is unclear. In summary, findings suggest it is difficult in the FO course context to optimize feedback timing to maximize understanding.

Overall, when participants’ accounts are examined with respect to these three steps, the nature of the FO context may be seen to present a challenging environment. One that suggests it is difficult ‘to hold [FO] learners in their ‘zone of proximal development’ by providing just enough help and guidance, but not too much’ (Perkins, 1992, p.163).

Turning to the use of CAA-provided feedback, based on findings from both thesis studies, this thesis argues that problems with student-instructor and student-student interactions in FO mathematics courses are leading to decreased use of discussion which is encouraging, as discussed in the second study results, increased dependence on CAA systems (i.e. student-technology interactions). Furthermore, limited to the extent that most participants appear to be using only one CAA system (i.e. MML), the quality of CAA-provided feedback is identified by interview participants as inferior to instructor-provided feedback and findings suggest a different set of issues that differentiate this feedback from instructor-provided feedback. In particular, identified issues with stage one CAA-provided initial feedback appear significantly amplified in follow up stage two feedback, where the assessment discussion is better characterized as an
interaction. The following discussion considers, this time separately, the three feedback steps associated with each stage. First, the three steps associated with participants’ stage one CAA-provided initial feedback are discussed. Second, the three steps associated with participants’ stage two CAA-provided assessment interactions, which is associated with iterative attempts and immediate feedback, are discussed.

With CAA-provided initial feedback evidence of student learning is identified as limited to only a single answer and not the progression of steps needed to solve a problem. As participants’ accounts suggest (e.g. P4: ‘to have the computer diagnose what the students [think]...that’s difficult to do on a computer’), it is difficult for CAA systems to accurately judging student thinking (Hrastinski & Stenbom, 2012; Skemp, 1976).

In the second feedback step, the selection of appropriate feedback, participants identify only generic not student-specific forms of feedback. While participants’ accounts suggest this feedback may be any one of the three types considered in this thesis, the lack of specificity, particularly with regards to addressing student thinking, places some doubt on its usefulness in informing student learning (Gibbs & Simpson, 2004). For example, there appears to be no identified ability to treat errors as ‘misconceptions that are actually preconceptions’ (Tall, 2012). Instead, the kind of feedback provided is more consistent with ‘directive’ rather than ‘responsive’ support, which research at the school level has linked with using digital technologies to develop students’ understanding mathematics (Walshaw, 2012).

In the third and final feedback step of timing feedback, as discussed in the literature review and found in this research, CAA-provided feedback is associated with immediate feedback which may be provided after an answer or a completed instrument is submitted. In the first initial feedback stage, this provision is viewed favourably by participants with time proximity of error detection to actual problem attempt as well as affective benefits identified (e.g. Shute, 2008). However, this appears to contrast with instructor-provided feedback where problem areas in student thinking may be ‘immediately apparent to the sensitive teacher’ (Tall, 1977, p.11) and feedback may be delayed depending on the kind of error (Kluger & DeNisi, 1996).

In the second follow-up stage of CAA-provided assessment interactions, participants associate this interactivity with iterative almost identical\(^{41}\) question or instrument attempts combined with

\(^{41}\) Most participants appear to characterize their use of CAA with successive attempts at identical or almost identical problems. Responses from two participants illustrate some variation between attempts. First, P2 stated that s/he...
immediate feedback. In relation to learning outcomes, findings indicate that half the participants express little concern about this interactivity with some noting, consistent with claims in the literature (e.g. Shute, 2008), that CAA systems are limited to helping students develop procedural understanding. The remaining participants express concern that gains in procedural understanding that are not flexible or deep (c.f. Baroody et al., 2007). Some go as far as suggesting (e.g. P5: '[CAA] mitigates against... real understanding’) that this interactivity is detrimental to developing higher-level understanding. This is consistent with what Simmons & Cope (1993) found, as discussed in the literature review, regarding the use of immediate feedback in school-level computer-mediated mathematics instruction. Perhaps more seriously, participants identify some students engaged with this interactivity as having arrived at the right answer but clearly the wrong understanding. Similar to what Sangwin et al. (2010) refers to as having an ‘automatic strategy [but] no underlying mathematical understanding’ (p.243).

Moreover, consistent with findings that ‘feedback system[s] [structure] students’ learning strategies’ (Whitelock and Raw, 2003, p.712), participants identify some students developing a dependence on CAA feedback.

In summary, in their experience of using CAA, limited by the suggested predominant use MML, initial feedback is generic, not student-specific, and is given immediately upon submission of a single answer. In addition, this feedback does not appear to specifically address a students’ thinking. In the second stage, participants associate the use of CAA with iterative attempts and immediate feedback. This kind of interactivity is identified with, at best, helping develop some level of procedural understanding and, at worst, being detrimental to the development of deeper understanding.

12.4.1 Limitations

The analysis suggests three limitations. First, the actual benefit to learning of instructor-provided feedback is limited by how that feedback is actually being used. As Gill and Greenhow (2008) conclude in their research: ‘real learning’ takes place ‘provided that [students] have truly engaged...especially by spending time studying the feedback provided’ (p.207; italics mine). With this in mind, these feedback findings are limited to the potential rather than actual
effect of this feedback. Second, participants are not asked to distinguish their use of initial and follow-up feedback. For this study, where it is not obvious, the present analysis infers when participants are discussing initial feedback or follow-up feedback. Further research should consider these stages separately. Third, as previously discussed, it appears most participants’ responses relate to their experiences with MML and, even if a greater variety of CAA systems were used, the functionality and nature of CAA systems and feedback is considered to be constantly shifting. For example, adaptive forms of CAA feedback tailored to individual students are emerging (Larreamendy-Joerns, Leinhardt, Corredor, 2005).

12.4.2 Summary

In support of feedback literacy findings that emerged in the first study, participants primarily emphasized how feedback is provided (i.e. process) despite attempts to direct participants to address the kind(s) of feedback they used. These findings suggest that FO instructors may be providing better initial feedback (e.g. more detailed feedback) in their FO than their F2F courses, but are finding it difficult to use this feedback in assessment discussions with their students. In other words, participants are struggling to move their feedback provision from stage one to stage two feedback. This appears linked with some participants’ identifying FO course feedback as ‘more summative’ (P4) than F2F feedback. These findings are consistent with what was previously discussed about participants’ struggles with using discussion in general. Furthermore, they present more evidence (e.g. encouragement to use live office hours in FO courses and using quicker feedback) that FO instructors are struggling to provide the reflective interactivity needed in mathematics instruction.

Finally, regarding the use of CAA, where the use of MML is predominant, findings suggest that feedback is not-specific to student thinking and that the two-way feedback interaction between CAA and the student at best benefits some level of procedural understanding but may be detrimental to developing deeper conceptual understanding.

12.5 How are participants’ approaching their FO course assessment practice?

Findings from the first study did not suggest, as expected from the literature (e.g. Nelson Laird et al., 2008), an overall orientation to KR (see Table 34). This raised questions about some possible influence from an emphasis on constructivist pedagogy in FO instruction (Anderson & Elloumi, 2008).
Table 34: Overall summary of participants' orientations

<table>
<thead>
<tr>
<th>Orientations to assessment practice</th>
<th>Belief Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nature &amp; structure of knowledge</td>
<td>2. Degree of integration of knowledge - Q11</td>
</tr>
<tr>
<td>3. Degree of transformation of knowledge - Q12</td>
<td>4. Differences between good &amp; poor answers - Q13</td>
</tr>
<tr>
<td>5. Role of assessment in teaching &amp; learning – Q14</td>
<td>6. Use of feedback gained from assessment – Q15</td>
</tr>
<tr>
<td>KC</td>
<td>unclear</td>
</tr>
</tbody>
</table>

Table 35: Orientation of participants’ F2F vs. FO assessment schemes as identified by total assessment weightings (with weighting associated with the use of CAA specified)

<table>
<thead>
<tr>
<th>Participant</th>
<th>F2F Course</th>
<th>FO Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KR</td>
<td>mid</td>
</tr>
<tr>
<td>P1</td>
<td>100% CAA-based</td>
<td>0</td>
</tr>
<tr>
<td>P2</td>
<td>80% (with 20% CAA-based)</td>
<td>5%</td>
</tr>
<tr>
<td>P4</td>
<td>8% CAA-based</td>
<td>85%</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>20%</td>
</tr>
<tr>
<td>P6</td>
<td>16.7%</td>
<td>50%</td>
</tr>
</tbody>
</table>

In this second study, further insight into approaches to assessment in FO courses was sought by asking each participant how each component of their FO and F2F assessment schemes was oriented according to S&B’s orientations (with three possible choices: KR, KC or ‘midway’; see Appendix M for detailed results) and then totalling up the associated weightings. As a result, compared to their F2F course, two of five participants’ FO course assessment instrument weightings are found to be clearly more associated with KR (P4 and P5), two only slightly more associated with KC (P2 and P6) and one the same in both contexts (P1; see Table 35).

In contrast with the first study, the second study suggests participants’ FO courses are more oriented to KR than their F2F courses. These differences will now be discussed by considering the differences in F2F and FO course assessment weightings associated with KR, KC and midway.

**Participants Whose FO Course Assessment Weightings Are More Oriented to KC**

Two participants identify their FO course assessment schemes and weightings as only slightly more oriented to KC. First, P2’s FO course is only 5% more in favour of KC while 5% less in favour of midway, with no difference in KR associated weightings. It is inferred from P2’s

42 P3 was not included because the F2F and FO courses were not the same.
responses that this is due to the identification, consistent with some of the literature (Offenholley, 2006), of FO ‘discussion’ as more oriented to KC than F2F ‘class participation’, which is considered to have a midway orientation. As P2 states, for his/her FO discussion students are being asked to do ‘some very specific things...[with] the discussion boards’. For example, every student is required to post ‘at least three times for each chapter’ where one post is what each student ‘thought was the most or least helpful in understanding the material in that chapter’. Second, P6’s FO course is only 5.5% more in favour of KC and 1.7% more in favour of midway while 8.5% less in favour of KR. This is due to P6’s use of additional instruments in his/her FO course (i.e. ‘written assignments’, ‘exploratory assignments’ and ‘virtual class’), all identified as more oriented to KC. This is consistent with claims in the literature (Gikandi et al., 2011) as well as the first study, where the use of a greater variety of instruments was found to be associated with higher S&B measures. As P6 states, extra assessment instruments are used to mediate for the lack of ‘intuitive assessment that happens when [s/he is] sitting next to them’ so as to make sure his/her students ‘really are getting’ the mathematics.

Participants Whose FO Course Assessment Weightings Are More Oriented to KR

Next, two participants identify their FO course assessment schemes and weightings as considerably more oriented to KR. First, P4’s FO course is 79.5% more in favour of KR, 80.5% less in favour of midway and 1% more in favour of KC. This is primarily due to P4’s heavy reliance on CAA (and specifically MML) in his/her FO course and that s/he identifies the use of CAA with a KR orientation and the use of paper-based instruments with a midway orientation. As P4 states ‘it’s very clear’ that the ‘fault’ is with the ‘computer system’ which is ‘clearly reproducing and regurgitating the skills and the manipulative skills’ and cannot ‘test conceptual understanding or, [the] ability to problem solve’. Second, P5’s FO course, with no difference in KR associated weightings, is 45% more in favour of midway but 45% less in favour of KC. This is primarily due to differences in P5’s invigilated exams where s/he identifies his/her FO course exams as more oriented to midway and his/her F2F course exams as more oriented to KC. As P5 states, because s/he is having ‘difficulty preparing [FO] students for open-ended questions’ s/he feels less ‘confident in...[asking] those sorts of questions [on an exam]’. Instead, s/he tends to emphasize ‘procedural learning in... [FO] exams’. Finally, s/he links this difficulty with a lack ‘real time’ interactions that permit opportunities for collaboration and scaffolding.

This question investigated participants’ views about how each of their course assessment instruments are oriented relative to the weighting it receives in the course assessment schemes.
The overall difference in participants’ identified orientations in F2F and FO course contexts is that when an FO course is considered more oriented to KC this is only slight but when it is considered more oriented to KR the difference is considerable.

12.5.1 Limitations

One possible limitation, applicable to much of the present research, concerns the nature of the virtual learning environment(s) (VLE) used by study participants. Because the study made no attempt to investigate the VLE each participant was using, it is unclear what impact, for example, the VLE design had on participants’ attempts to orient their assessment practices to KC (Mueller & Strohmeier, 2011). For future research, it would be valuable to consider the nature of the VLE used by participants.

12.5.2 Summary

These findings present a somewhat complex picture of approaches to assessment practice in current FO mathematics courses. The first study seems to suggest instructors’ FO course orientation to assessment may be more oriented to KC than F2F courses. However, the second study, where five of the six US public HE participants’ F2F and FO course approaches to assessment are directly compared, suggests that their FO courses are overall less oriented to KC than their F2F courses.

One possible explanation for this disparity, consistent with prior findings (Murray & Macdonald, 1997), may be that participants, when answering the first study S&B survey questions, may be interpreting ‘approach’ in terms of intention or belief and not actual practice. Two interview participants’ comments, left along with their S&B question responses, are consistent with this explanation. First, P6, leaving a comment with his/her answer to the fifth S&B question (Q14), states ‘the answer above reflects what I believe about assessment...[but not what is actually happening in practice]’. On the other hand, reflecting some of the tension participants may be experiencing in answering the S&B questions, P4, leaving a comment with the third question S&B question (Q12), states ‘I'm almost embarassed [sic] to admit that... ’.

However, on balance, the findings for this question agree with participants’ accounts on the use of discussion, quizzes and feedback where the FO course context is revealed to be a challenging environment for developing students understanding of mathematics. For example, despite what may seem like better initial assessment feedback in FO courses, assessment discussions that
could help facilitate closing the gap between students’ present and desired state of learning (Hattie & Timperley, 2007) appear restricted and some participants are encouraging students to use their F2F on campus office hours for discussion. Also, participants describe having much less flexibility in their FO than their F2F courses, where, for example, quizzes may be used collaboratively to realize KC. In summary, findings suggest that participants are struggling to orient their assessment practice to KC in their FO courses. Moreover, as previously discussed, where CAA systems such as MML are identified as favouring orientations to KR and there is a greater dependence on these systems in FO than F2F courses, this would further suggest participants’ approaches to assessment practice in their FO as compared to their F2F courses are less oriented to KC.

12.6 Summary of Research Findings

Central to the mixed methods approach used in this thesis, the second study used qualitative research methods to explain some of the first study findings and gain more insight into current teaching and assessment practices of FO mathematics instructors’ that teach in US public HE institutions. The results of this investigation are summarized below for each research question.

*Study II - R1: How and why is discussion/interaction used?*

A minority (39%) of survey participants used discussion as a weighted assessment instrument. Providing an explanation for this low level of discussion use, all interview participants report problems conducting discussions in their FO courses, problems that have led some to opt out of emphasizing or even encouraging any student-student interactions. This is consistent with the balance of evidence discussed in the literature review that discussion is not working well in FO mathematics courses. Analyses of interview responses suggest two root issues. Specifically, the current nature of the FO course context is not compatible with the disciplinary nature of reflective interactivity and communication resources needed for effective mathematics instruction. In terms of reflective interactivity, participants identify real-time interactions as necessary for effective instruction. In terms of communication resources, they are limited to text alone, with no facial expressions visible nor voice tones or pauses that can be used or heard. As a result, participants are struggling to judge the state of student thinking as well as communicate their own thinking.
**Study II - R2: How and why are quizzes being used?**

About half (53%) of survey participants used quizzes as a weighted assessment instrument, fewer participants than expected based on the literature (Engelbrecht & Harding, 2005a; Greenberg & Williams, 2008). Most of these are online quizzes (81%) that provide immediate feedback. Analyses of interview responses suggest these quizzes are associated with the provision of repeated attempts and used primarily to direct rather than advance student learning. Where any learning occurs, they are seen to, at best, aid the development of procedural knowledge and, at worst, act as a detriment to the development of deeper conceptual knowledge. This is in stark contrast to F2F instructor-administered quizzes which interview participants identify as a much more potent tool (e.g. allowing for impromptu collaborative use) for advancing student learning.

**Study II - R3: How and why are participants choosing to use invigilation?**

Analyses of interview responses suggest one primary purpose for which invigilation is used – the validation of student learning – that is consistent with prior claims (Trenholm, 2007b, Flesch & Ostler, 2011). In addition, where traditional F2F invigilation practice appears largely assumed and unchallenged, it is relatively easy for FO mathematics instructors to choose not to use invigilation given, for example, pressure from administrators combined with the potentially onerous time requirements for administering invigilated exams to remote students.

Finally, one interview participant reported better quality learning when using a combination of invigilation and non-invigilation with one summative-style assessment instrument. However, as reported in the first study, overall analyses of survey participant data appear to reveal contradictory findings. That is, despite providing richer feedback (e.g. Hattie & Timperley, 2007) than those using invigilation, these participants were more likely to orient their assessment practice to KR than those using invigilation. One possible reason, suggested by the second study findings, is that this feedback is mostly associated with intermediate feedback goals identified by participants, goals that are not necessarily directly linked to advancing student learning.

**Study II - R4: How and why is feedback being provided?**

Analyses of survey participant data suggested possible problems with feedback literacy. This suggestion appears to be supported in the interviews when analyses of participants’ responses suggested they connect the quality of their feedback practice primarily with *how* and not – as
emphasized in the literature (e.g. Hattie & Timperley, 2007) – with what kind of feedback is provided. In their FO courses, while initial feedback may be better than what is given F2F (i.e. quicker and more detailed), findings suggest it is difficult to leverage this feedback for use in assessment discussions (e.g. Dunn & Mulvenon, 2009), which have the potential to further learning. This suggests that the nature of assessment in FO courses would be ‘more summative’ (P4) and more individualized, making less use of collaboration. Finally, interview participants suggest the kind of feedback interactivity provided by CAA systems is leading to the development of procedural knowledge that is not deep and may even be conditioning students to be dependent on CAA feedback, thereby gaining a false sense of understanding. Findings related to CAA are limited to the extent most participants appear to be using the same CAA system, MML.

**Study II - R5: How are participants’ approaching their FO course assessment practice?**

Survey participants, in the context of FO mathematics assessment practice, appear to be less oriented to KR than expected (i.e. for F2F instruction, e.g. Nelson Laird et al., 2008). Analyses of interview participant responses suggest otherwise, that their FO mathematics courses are more oriented to KR than their counterpart F2F courses. As the second study findings further suggest, environmental constraints presented by the FO course context (as well as CAA systems, limited by the predominance of MML) favour orientations to KR and make it difficult for participants to orient their assessment practice to KC.
13. Conclusions

This research sought to investigate the nature of current FO mathematics courses through the lens of instructors’ assessment practices. A mixed methods research approach was employed. In the first study, FO mathematics instructors, who were mostly from the US, were surveyed to explore their current assessment practices and approaches. In the second study, semi-structured interviews were conducted with a selection of the first study participants to gain further insight into particular practices identified by the first study and the literature. All second study participants taught in US public HE institutions, where FO instruction is known to be prevalent, and together represented a range of approaches to teaching and assessment (as measured in the first study).

The two studies appeared to present contrasting findings regarding instructors’ approaches to assessment in their FO courses. In the first study, no clear overall orientation to KR was found among participants. This was in contrast with expectations from prior studies on F2F tertiary mathematics instruction which suggested participants’ approaches to assessment would be clearly oriented to KR. In the second study, participants’ accounts suggested that, overall, it was much more difficult to develop students’ understanding of mathematics in the FO as compared to the F2F course context. Moreover, when directly comparing participants’ F2F and FO course assessment schemes, the second study suggested that FO courses are more oriented to KR than F2F courses. One possible explanation for these contrasting findings is that participants’ responses to the S&B questions in the survey reflected their beliefs or intentions, not their actual practice.

Helping to explain why FO courses are more oriented to KR than F2F courses, the second study analysis revealed two root issues related to the asynchronous nature and limited communication resources that typify current FO course contexts. That is, these contexts do not appear to be permitting the needed reflective interactivity or providing the needed communication resources identified as fundamental to how mathematics is instructed assessed and ultimately learned for depth of understanding. For example, findings suggest it is more challenging to judge, select and time feedback for the purpose of developing student understanding in FO courses. Perhaps understandably, participants’ accounts of adaptation revealed a struggle to mediate for these problems. For example, in FO as compared to F2F courses, quizzes appeared to be used more
for the purpose of directing student learning and feedback appeared to be used more for the purpose of encouraging self-assessment and enabling interactions.

These issues suggest a possible reason why a de-emphasis on the use of discussion as a weighted assessment instrument was found in the SUNY context in the first study, when comparing FO courses over several years. And consistent with this de-emphasis, they also provide a reason why prior research has found hard pure FO courses becoming less oriented to community practice (Smith, Heindel et al., 2008) and FO mathematics courses, in particular, a more ‘one-on-one’ experience with little use of collaboration (Lowrie & Jorgensen, 2012). In balance, based on the thesis participants’ practices and experiences, this research suggests problems with human interactions (i.e. student-student and student-teacher interactions) in FO mathematics courses.

In addition, both studies provide evidence that suggests instructors are more reliant on CAA systems for their FO courses than their counterpart F2F courses. Participants did not identify CAA systems with the kind of reflective interactivity deemed necessary for helping to develop students’ understanding of mathematics. Instead, the learning resultant from CAA-provided interactivity was viewed by participants to, at best, aid in lower-level procedural thinking and, at worst, be detrimental to the development of students’ mathematical thinking. However, apart from evidence of a greater dependence on CAA systems, participants’ views of these systems are limited to the extent most used the same system (i.e. MyMathLab).

Changes in the nature of interactions are of interest given they are considered key to student learning in FO courses (Moore, 2002; Anderson & Elloumi, 2008) as well as distance education (Moore, 2007). In relation to current theory, problems with human interactions in general (including a de-emphasis on student-student interactions) and an increased dependence on the use of CAA present a dynamic consistent with distance education theory (Moore, 2007). That is, as this research appears to suggest, current FO mathematics instruction is evolving such that problems with human interactions are leading to a greater dependence on student-content\textsuperscript{43} interactions.

However, the nature of FO courses, as reflected across participants’ accounts of teaching and assessment in their FO courses, suggests a degraded learning experience. Responses from both

\textsuperscript{43} Some refer to student-technology interactions as an additional category (Vrasidas, 2000). Here, student-content is considered to include student-technology interactions.
community college instructors (P5 and P6), neither using CAA, appear to reflect the balance of evidence concerning the quality of current FO mathematics courses:

P5: But I certainly believe that online courses can be effective, and need to be effective, but from my perspective, the biggest problem is lack... of focusing on concepts and rather simply focusing on procedures.

P6: I would have to be honest and tell you that I don't think the online course has the quality that my in-class course has... I have racked my brain...

In addition, the identified nature of CAA-based learning, though largely pertaining to the use of MML, suggests the use of CAA contributes to making FO mathematics courses more oriented to KR. A comment left by P4, appears to sum up the balance of evidence concerning the dependence on CAA, and specifically MML, in FO mathematics courses:

P4: Using the software product that we do... [t]he level of conceptual teaching and assessment is much lower in my online class than in my F2F class, or at least I'm less able to determine the degree to which students really understand the material.

Finally, these outcomes, considered alongside changes in interactions, appear to challenge current theory in online learning. Anderson (2003) theorizes that ‘deep and meaningful learning can be developed as long as one of the three forms of interaction (student–teacher; student–student; student–content) is at very high levels while the other two may be offered at minimal levels or even eliminated without degrading the educational experience’ (p.4). However, participants’ accounts suggest that deep and meaningful learning is not being realized even though evidence suggests human interactions are at lower levels while student-content interactions are at higher levels. As suggested by the second study, what may be needed for deep and meaningful learning is to address the nature and not just the level of interaction.

13.1 Implications for Practice

Previous research suggests that changing assessment practice alone is not enough to realize deep learning in students (Crawford, Gordon, Nicholas, & Prosser, 1998; Marton & Säljö, 1984). While the present research focuses on assessment practice it is clear it has implications for

44 Left with the third S&B question on the ‘Degree of Transformation of Knowledge’, which was the twelfth question on the survey.
general instruction as well (e.g. use of discussion). Two principal implications for practice are recommended:

1. Where current pedagogical research and development work has generally failed to address disciplinary characteristics, this research demonstrates the need for mathematics-specific FO course development. This need is particularly identified regarding the necessity of F2F synchronous interaction for teaching and assessing FO mathematics courses for depth of understanding. In short, specific attention needs to be directed at instructional design of FO mathematics courses, particularly related to the nature of interactions they afford. Moreover, in this respect, professional development activities need to be focused on addressing the particularities of mathematics pedagogy in this environment.

2. Though limited by the predominant use of MML, many questions are raised about the quality of learning produced by using CAA systems. These questions suggest there is considerable room for improvement. The use of CAA appears to be both complicated and potentiated by ongoing technological developments. Ongoing professional development is needed for instructors and course developers to keep abreast of current research and technological developments. Where CAA systems may ‘promise the moon’, such work needs to help instructors separate the rhetoric from the evidence (Lynch, 2006, p.32).

13.2 Recommendations for Further Research

The present research investigates instructors’ experiences of their teaching and assessment practice. Apart from, for example, attempting to replicate these findings, many of these recommendations (appearing in the form of additional research questions) are directed at investigating students’ experiences of FO as compared to F2F learning:

1. The ATI and S&B findings are based on interviews from academics outside of mathematics. How would these findings be different if they were only based on responses from mathematics instructors?

2. In the context of mathematics instruction, what is the student experience of reflective interactivity and uni-modal vs. multi-modal communication?

3. Evidence suggests a greater reliance on CAA in FO as compared to F2F mathematics courses. Can these findings be replicated with a direct investigation of the use of CAA
in FO compared to F2F courses? In particular, can any new findings confirm a direct relationship between the use of CAA and the status of student-student and student-teacher interactions in FO as compared to F2F mathematics courses?

4. Participants’ views concerning the use of CAA appear limited by the predominant use of MML. How are these views representative of CAA systems, in general, or just MML?

5. The interview participants identify different ways they perceive and discover their students to be engaging with CAA. From the student perspective, how are they actually engaging with CAA? Does the use of CAA, for example, ‘interfere’ with students attaining deeper understanding (Entwistle, 2009, p.85) or is it acting as a gateway to deeper understanding (Baroody et al., 2007)?

6. The research raises questions regarding the way instructors are going about providing feedback (i.e. possible issues of feedback illiteracy). Do these questions persist when the use of the study measures are expanded to a different and larger population of mathematics instructors?

7. Is there a relationship between assessment weighting and how students engage with the associated instructor feedback?

8. The thesis findings suggest it may not be possible to teach and assess mathematics for depth of understanding in the current FO course context. Does this mean that students cannot learn mathematics for depth of understanding in FO mathematics courses as well as they do in a F2F course? Is any difference based on demographic characteristics, as some prior research suggests (e.g. McIntosh & Morrison, 1974)? Further research should investigate how students experience their learning in FO mathematics courses compared to the same or similar F2F courses.

9. The use of invigilation is expected in F2F but not necessarily in FO mathematics courses. How are students engaging in learning mathematics in non-invigilated FO mathematics courses? What is the relationship between students’ approaches to studying (e.g. Entwistle & Ramsden, 1983) and students preferring FO to F2F mathematics courses?

10. Is there a benefit to student learning when a single assessment instrument administered multiple times is given in a combination of invigilated and non-invigilated settings? Where assessment questions are, for example, more challenging or conceptual in nature in the non-invigilated setting?

11. What is the relationship between approaches to assessment or teaching and the nature of the virtual learning environment(s) (VLE) used by participants?
13.3 Concluding Statement

Some have hypothesized that online courses will transform the way mathematics is taught and learned (Borba, 2005). This thesis study, of mostly US FO mathematics instructors, suggests this transformation may currently be hindering the development of students’ understanding of mathematics and is more identified with a degraded learning experience. This is consistent with meta-analytic findings that suggest mathematics instruction appears ‘best suited to the classroom’ (Bernard, Abrami, Lou, Borokhovski, Wade et al., 2004, p.400) as well as prior findings, regarding FO mathematics instruction, on student satisfaction (e.g. Summers et al., 2005) and attrition (e.g. Mensch, 2010).

To address this issue, this research suggests FO development efforts need to be directed at the nature of interactions in FO mathematics courses and how to provide instructional components that include, for example, virtual F2F synchronous interactions. Harman and Dorman (1998) make a similar argument, suggesting that ‘videoconferencing and audiographics’ be used to ‘enrich’ distance teaching and learning of mathematics. However, beyond enrichment, this research, based on the practices of mostly US HE mathematics instructors and experiences of US public HE mathematics instructors, suggests that addressing the nature of interactions may be in fact fundamental to teaching and assessing for depth of understanding in FO mathematics courses.
References


Appendices

Appendix A

The original ATI questions (Trigwell & Prosser, 2004) organized according to the two scales with item numbers as they appear ordered in the survey. Responses to all questions are on a 5-point likert scale from ‘only rarely true’ (i.e. score of 1) to ‘almost always true’ (i.e. score of 5).

<table>
<thead>
<tr>
<th>Scale: Conceptual change/student-focused (CCSF) approach</th>
<th>Subscale: Conceptual change intention items</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATI05</td>
<td>I feel that the assessment in this subject should be an opportunity for students to reveal their changed conceptual understanding of the subject</td>
</tr>
<tr>
<td>ATI08</td>
<td>I encourage students to restructure their existing knowledge in terms of the new way of thinking about the subject that they will develop</td>
</tr>
<tr>
<td>ATI15</td>
<td>I feel that it is better for students in this subject to generate their own notes rather than always copy mine</td>
</tr>
<tr>
<td>ATI16</td>
<td>I feel a lot of teaching time in this subject should be used to question students’ ideas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subscale: Student-focused strategy items</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATI03</td>
</tr>
<tr>
<td>ATI06</td>
</tr>
<tr>
<td>ATI09</td>
</tr>
<tr>
<td>ATI14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale: Information transmission/teacher-focused (ITTF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscale: Information transmission intention items</td>
</tr>
<tr>
<td>ATI02</td>
</tr>
<tr>
<td>ATI04</td>
</tr>
<tr>
<td>ATI11</td>
</tr>
<tr>
<td>ATI13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subscale: Teacher-focused strategy items</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATI01</td>
</tr>
<tr>
<td>ATI07</td>
</tr>
<tr>
<td>ATI10</td>
</tr>
<tr>
<td>ATI12</td>
</tr>
</tbody>
</table>
Appendix B

S&B survey questions with possible responses based on S&B study framework.

1. **In your approach to assessment**, which of the following descriptions best describes how you view the nature and structure of mathematical knowledge?

   a. It is external to students and is a coherent body of knowledge structured by experts in the field.

   b. It is external to students and is a collection of important definitions, concepts, techniques, methods and theories.

   c. It is something that is internalised, reorganised, and reconstituted in the process of learning.

2. **In a typical assessment question**, which of the following descriptions best describes how you assess your students?

   a. Students integrate information from many sources, from more than one subject, and/or from their own experience.

   b. Students draw on information presented in a single lecture, tutorial, practical session or chapter.

   c. Students draw on information presented in many sources, but all within the course/module.

3. **In your approach to assessment**, which of the following descriptions best describes how you assess your students?

   a. Students apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems.

   b. Students apply well known techniques, methods, laws, principles, or explanations to unseen standard problems.

   c. I assess students on whether they can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes.

4. **In your approach to assessment**, which of the following descriptions best describes how you view the difference between good and poor answers?

   a. The difference lies in the accuracy and relevance of what is reproduced.

   b. Good answers are purposeful and justify the content used, whereas poor answers do neither of these things.
c. The difference between good and poor answers lies in the quantity of content correctly recalled.

5. **In your approach to assessment**, which of the following descriptions best describes how you use assessment in teaching and learning?

   a. Students have to be forced to study, and I use assessment as the best tool to achieve this.

   b. Assessment is an integral part of teaching and learning, a means of helping students learn.

   c. Assessment forces students to study, and marks give them an indication of the progress made and reward their efforts.

6. **In your approach to assessment**, which of the following descriptions best describes how you use feedback gained from assessment?

   a. I use feedback from student performance to monitor students' learning and to help them improve.

   b. I use feedback from student performance as a means of altering or adjusting my teaching approach.

   c. I use feedback from student performance to change my own or my students' actions.

   d. I use feedback from student performance to challenge students' existing ideas and understandings.
Appendix C

Screen shots of survey questionnaire as presented on the web for SUNY participants

---

**SUNY - Teaching Undergraduate Mathematics Fully Online: Assessment Approaches**

My name is Sven Trenholm and I am conducting research into how academics approach assessment in fully online* mathematics courses/modules. Your participation in this research will help support pedagogical and professional development in this growing area.

This international survey captures factual information about current assessment practices. The survey is completed anonymously, can be saved part way through and takes around 15-20 minutes to complete. The majority of the questions are multiple choice with opportunities for optional comments. Please answer all questions as honestly as you can as they relate to your actual practice.

Thank you very much for participating.

Sven Trenholm
Mathematics Education Centre
Loughborough University
Loughborough, Leicestershire
LE11 3TQ, UK
S.Trenholm@lboro.ac.uk
mathsresearching@gmail.com
+44 (0) 1509 22 8212

---

* Fully online courses/modules of interest to this survey are also referred to as fully "asynchronous" online courses/modules. These courses/modules generally have no requirement for attending actual physical lectures/classes/meetings - with the possible exception of exam invigilation and/or induction/orientation sessions.

---

**SUNY - Teaching Undergraduate Mathematics Fully Online: Assessment Approaches**

Data Protection

All data collected in this survey will be held anonymously and securely. No personal data is asked for or retained. Aggregate data may be retained to benchmark future surveys.

---

313
Demographic Questions (Section 1 of 4)

Please note: once you click on CONTINUE you will not be able to return to this page.

1. For how many years have you taught mathematics at the higher education level in a traditional face-to-face on-campus format?
   - 0
   - < 1
   - 1-5
   - 6-10
   - 11-15
   - 16+

2. How many years have you taught mathematics in a fully online format with no face-to-face contact (with the exception of exam invigilation and/or induction/orientation)?
   - 0
   - < 1
   - 1-5
   - 6-10
   - 11-15
   - 16+

3. Which best describes how you spend your time in academia?
   - Mostly research (education focus)
   - Mostly research (pure mathematics/statistics focus)
   - Mostly teaching
   - About the same amount of research and teaching
   - Other (please specify):

4. Which of the following best describes the kind of institution where you teach your fully online mathematics course/modules?
   - Traditional "brick and mortar" bachelor granting (e.g., university, 4 year college)
   - Traditional "brick and mortar" non-bachelor granting (e.g., North American community college)
   - Open/Online university/college
   - Other (please specify):

5. What is your job status at this institution?
   - Full-time staff/faculty
   - Part-time staff/faculty
   - Retired

6. In the last 3 years, in order from most to least, what are the fully online mathematics courses/modules that you have generally taught?

<table>
<thead>
<tr>
<th>Fully Online Course/Module</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Fully Online Course/Module 1</td>
<td></td>
</tr>
<tr>
<td>b. Fully Online Course/Module 2</td>
<td></td>
</tr>
<tr>
<td>c. Fully Online Course/Module 3</td>
<td></td>
</tr>
<tr>
<td>d. Fully Online Course/Module 4</td>
<td></td>
</tr>
<tr>
<td>e. Fully Online Course/Module 5</td>
<td></td>
</tr>
</tbody>
</table>

7. With respect to teaching mathematics in a fully online format ONLY, please indicate if you have engaged in any of the following?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes</th>
<th>Available, but no</th>
<th>Not available</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Workshops outside the institution focused on fully online instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Sabbatical leave to focus in whole or in part on fully online instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Travel funds paid by institution to attend professional development in fully online instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Internal grants for research into fully online instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Internal workshops and/or seminars focused on fully online instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. With respect to teaching mathematics in a **fully online format ONLY**, please indicate if you have engaged in any of the following?

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Available, but no</th>
<th>Not available</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. If you answered 'yes' in the previous question, please estimate the amount of training you have received (e.g. 3 hrs, 2 days, 5 weeks...). (Optional)

---

**SUNY - Teaching Undergraduate Mathematics Fully Online: Assessment Approaches**

**Assessment Approaches for Specific Course/Module Context (Section 2 of 4)**

Please note – once you click on CONTINUE you will **not** be able to return to this page.

### Course/Module Context

9. For the remainder of this survey it is important that you choose **one** course that you currently teach **fully online**, preferably a calculus or statistics course/module. If you do not teach calculus or statistics, please specify the name of the course/content you will focus on under 'other'.

- [ ] Fully online pre-calculus
- [ ] Fully online calculus I, II, or III
- [ ] Fully online introductory statistics
- [ ] Fully online advanced statistics
- [ ] Other (please specify):

   a. Please specify the level at which this course is taught:
   - [ ] Undergraduate
   - [ ] Graduate
   - [ ] Post-graduate
   - [ ] Other (please specify):

   b. How many years have you taught this specific course/module?
   - [ ] 0
   - [ ] 1-5
   - [ ] 6-10
   - [ ] 11-15
   - [ ] 16+

---

### Course/Module Specific Questions

For questions 10 to 15 please:
1. Refer only to the course context you specified in question 9.
2. Choose one answer that best describes your overall approach to assessment.

If you feel none of the options are satisfactory, please pick the closest possible and feel free to comment in the space provided.
10. In your approach to assessment, which of the following descriptions best describes how you view the nature and structure of mathematical knowledge?

- It is external to students and is a coherent body of knowledge structured by experts in the field.
- It is external to students and is a collection of important definitions, concepts, techniques, methods and theories.
- It is something that is internalised, reorganised, and reconstituted in the process of learning.

Please feel free to expand on your selection. (Optional)

11. In a typical assessment question, which of the following descriptions best describes how you assess your students?

- Students integrate information from many sources, from more than one subject, and/or from their own experience.
- Students draw on information presented in a single lecture, tutorial, practical session or chapter.
- Students draw on information presented in many sources, but all within the course/module.

Please feel free to expand on your selection. (Optional)

12. In your approach to assessment, which of the following descriptions best describes how you assess your students?

- Students apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems.
- Students apply well known techniques, methods, laws, principles, or explanations to unseen standard problems.
- I assess students on whether they can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes.

Please feel free to expand on your selection. (Optional)

13. In your approach to assessment, which of the following descriptions best describes how you view the difference between good and poor answers?

- The difference lies in the accuracy and relevance of what is reproduced.
- Good answers are purposeful and justify the content used, whereas poor answers do neither of these things.
- The difference between good and poor answers lies in the quantity of content correctly recalled.

Please feel free to expand on your selection. (Optional)

14. In your approach to assessment, which of the following descriptions best describes how you use assessment in teaching and learning?

- Students have to be forced to study, and I use assessment as the best tool to achieve this.
- Assessment is an integral part of teaching and learning, a means of helping students learn.
- Assessment forces students to study, and marks give them an indication of the progress made and reward their efforts.

Please feel free to expand on your selection. (Optional)

15. In your approach to assessment, which of the following descriptions best describes how you use feedback gained from assessment?

- I use feedback from student performance to monitor students’ learning and to help them improve.
- I use feedback from student performance as a means of altering or adjusting my teaching approach.
- I use feedback from student performance to change my own or my students’ actions.
- I use feedback from student performance to challenge students’ existing ideas and understandings.

Please feel free to expand on your selection. (Optional)
### SUNY - Teaching Undergraduate Mathematics Fully Online: Assessment Approaches

#### Assessment Specifics for Specific Course/Module Context (Section 3 of 4)

Please note - once you click on CONTINUE you will not be able to return to this page.

**Course/Module Assessment Specifics**

**Type of Assessment**

<table>
<thead>
<tr>
<th>Type of Assessment</th>
<th>Weighting in Mark (check all that apply)</th>
<th>Number of Items (check all that apply)</th>
<th>Self - Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Assessment Component 4</td>
<td>Direct or write</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Assessment Component 5</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Assessment Component 6</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Assessment Component 7</td>
<td>Select or answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 5 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 6 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 7 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 8 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 9 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Visible Feedback for Questions 10 Course/Module Context:**

**Typical timing to feedback (in minutes, 5 min, 24 hrs.):**

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Timing to Feedback</th>
<th>Self</th>
<th>Group (e.g., department)</th>
<th>Comprehension (e.g., textbook)</th>
<th>Multiple choice</th>
<th>Short answer</th>
<th>Long answer</th>
<th>Infused</th>
<th>Not infused</th>
<th>Weighting</th>
<th>Number</th>
<th>Infused</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Assessment Component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Assessment Component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Assessment Component 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Please feel free to comment on your assessment specific and/or feedback.**

(Typewriter)
Appendix D

Sample initial email for potential online survey participants:

To Whom It May Concern,

I am researcher in the UK at Loughborough University’s Mathematics Education Centre and I am writing to very kindly ask for your help.

I am conducting research using a confidential and anonymous online survey (link: https://www.survey.lboro.ac.uk/mathelearning) to investigate faculty approaches to assessment in fully online mathematics courses.

If you can spare a short period of your time (about 15-20 minutes) in the next week or so, your participation in this research will help further fully online mathematics pedagogy. Also, if you are interested, I will gladly share the results of the survey when all responses have been received (there will be a space to request this at the end of the survey).

If you have any questions please feel free to email me.

Thank you very much for your help.

Sincerely,

Sven (Trenholm)
Appendix E

Problems with the participant data with the actions taken are detailed in the table below.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Action taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>No academic level given for Mathematica for calculus, Mathematica for</td>
<td>All are considered as undergraduate courses given last course was taught at a two-year college and others are assumed undergraduate based on the course title.</td>
</tr>
<tr>
<td>pre-calculus and Nature of Mathematics</td>
<td></td>
</tr>
<tr>
<td>BEng Mechatronics/Quality Management/Electronics/Polymer Engineering</td>
<td>It is assumed this participant misunderstood the question. This data is excluded from previous course history.</td>
</tr>
<tr>
<td>was listed as ‘modules’ at the Institute of Technology of Ireland.</td>
<td></td>
</tr>
<tr>
<td>Total weighting adds to 99%</td>
<td>Is left at 99%.</td>
</tr>
<tr>
<td>Total weighting adds to 70% with ‘0’ inputted for homework weight</td>
<td>It is assumed homework weight was intended to be inputted as 30%. Homework weight is changed to 30%.</td>
</tr>
<tr>
<td>Total weighting adds to 80% with ‘2’ inputted for final exam weight</td>
<td>It is assumed final exam weight was intended to be inputted as 20%. Final exam weight is changed to 20%.</td>
</tr>
<tr>
<td>Total weighting adds to 100% with homework assigned and nothing inputted</td>
<td>It is assumed homework weight was 0%. Homework weight is changed to 0%.</td>
</tr>
<tr>
<td>for its weighting</td>
<td></td>
</tr>
<tr>
<td>Total weighting adds to 95%</td>
<td>Is left at 95%.</td>
</tr>
<tr>
<td>Total weighting adds to 110%</td>
<td>It is assumed it is a bonus. Is left at 110%.</td>
</tr>
<tr>
<td>‘NA’ or nothing inputted for final exam feedback (2 participants)</td>
<td>It is assumed no feedback is given apart from the grade.</td>
</tr>
<tr>
<td>Nothing inputted for discussion feedback (2 participants)</td>
<td>It is assumed no feedback is given apart from the grade.</td>
</tr>
<tr>
<td>Nothing inputted for discussion and group project feedback (1 participant)</td>
<td>This data is excluded from the analysis.</td>
</tr>
<tr>
<td>‘Other’ assessment instrument category is used three times by one</td>
<td>Given feedback characteristics are identical, they are considered as a single instrument in the analysis related to feedback. However, when considering the variety of assessment instruments, they are considered as three separate instruments.</td>
</tr>
<tr>
<td>participant</td>
<td></td>
</tr>
<tr>
<td>‘Homework’ category used more than once (2 participants)</td>
<td>Second homework instrument is listed under ‘other’ category given assessment specifics are different than other homework instrument listed.</td>
</tr>
<tr>
<td>‘Quiz’ category used more than once (1 participant)</td>
<td>Second quiz instrument is listed under ‘other’ category given assessment specifics are different than other quiz instrument listed.</td>
</tr>
<tr>
<td>Number of times (i.e. frequency) ‘final exam’ is given missing (12</td>
<td>It is assumed the participant administered only one final exam.</td>
</tr>
<tr>
<td>participants)</td>
<td></td>
</tr>
<tr>
<td>Number of times (i.e. frequency) ‘mid-term exam’ is given missing (6</td>
<td>It is assumed the participant administered only one mid-term exam.</td>
</tr>
<tr>
<td>participants)</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix F

### Q18 – Please feel free to comment on your assessment specifics and/or feedback (Optional).

<table>
<thead>
<tr>
<th>Original Uncorrected Comments (in no specific order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• One type of feedback seems to be missing: feedback that aims at the regulation of the learning process. This means “Hints or comments that support the learner in the selection of content and the regualtion of learning.” I see this as important feedback that is different from &quot;Hints or comments challenge student understanding &quot;!</td>
</tr>
<tr>
<td>• online teaching can work very well for motivated students if the focus is on discussion rather than solving exercises. answers to most problems are available on sites such as <a href="http://www.wolframalpha.com/">http://www.wolframalpha.com/</a> or for little money per question at other sites (sites such as <a href="http://www.wolframalpha.com/">http://www.wolframalpha.com/</a> are good for society!!)</td>
</tr>
<tr>
<td>• During the homework sessions students can view solutions to similar problems and received hints and help, but solutions to quizzes and exams are only available after the completion of the assessment.</td>
</tr>
<tr>
<td>• The grade is given immediately after the attempt but the rest of the feedback such as correct answer and lecturer comment is later after the quiz has closed. also I do the assessment in class after it is closed.</td>
</tr>
<tr>
<td>• The individual projects are Minitab assignments where interpretation of results is emphasized. For the discussions, I read and comment within 24 hours. The discussion lasts for 2 weeks before grades are given.</td>
</tr>
<tr>
<td>• We are going to allow proctored tests/exams beginning next fall. The decision will be left to the instructor and I would choose to proctor Final Exams.</td>
</tr>
<tr>
<td>• Homework and quizzes can be attempted 3 times and the highest grade is used in calculating the final grade.</td>
</tr>
<tr>
<td>• For exams some parts they get immediate feedback other parts when they are manually graded.</td>
</tr>
<tr>
<td>• The midterm and Final are taken on campus while the other tests are taken at home and sent in the mail or faxed. A student MUST have a passing average on thh midterm and the fnal or they will not pass the course. I use MathXL to generate the homework and the online Quizzes.</td>
</tr>
<tr>
<td>• Question 17 does not exactly reflect what I do. The project is graded only +/0/-, where those indicate i/2 letter grade changes in the final course grade. A student who submits nothing loses a full letter grade. The effect approximates what I have written. The project is the last element of the course, so students (should) have adequate time to do well.</td>
</tr>
<tr>
<td>• Midterm exams are scanned into the computer and videos are made of the grading process. Students then receive a video along with an ungraded PDF. They can write corrections as they watch their video.</td>
</tr>
<tr>
<td>• 5% is assigned as a “participation” score - this is used as leverage to get students to log in on a regular basis, submit assignments on time, participate in informal online discussions with their group, etc. e) consists of assignments in which students have to select a problem from specified section that relates to their career interests, and write a short paper related to that problem - in one of a variety of specified formats (eg a letter to your aunt explaining how your work in calculus is going to be relevant to your career, or a letter to a newspaper on the topic, etc.)</td>
</tr>
<tr>
<td>• component 3: individual lab experience (using minitab) ten total for the term, + 4 projects where the student gathers data and illustrates the various techniques used in the labs.</td>
</tr>
<tr>
<td>• Turnaround time on feedback is determined by number of papers to be marked, number of markers available, extensions of time given to students due to illness. No feedback other than final mark is given for final examination.</td>
</tr>
<tr>
<td>• Also applied is a rubric. Three components 1. Basic skills via ALEKS work and quizzes 2. Conceptual ideas via LMS lecturer authored quizzes and activities 3. Final exam (all multiple choice) Above, I based my “weight” on 200 points each. However, I also weight the Lowest of the above at 200 points (counts twice) and the highest of the above at 200 points (counts twice).</td>
</tr>
</tbody>
</table>

### Q20 – Please feel [sic] to comment on any aspect of this inventory. (Optional)

<table>
<thead>
<tr>
<th>Original Uncorrected Comments (in no specific order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Question &quot;O&quot; is not appropriate when course materials are provided.</td>
</tr>
<tr>
<td>• I believe that an instructor should know the material they are teaching but I enjoy challenging questions from my students that I might not know the answer because then I can turn it around for a discussion. I feel that students, given various materials, should be able to critical think about what they are learning and should be able to draw their own inferences and conclusions based on the evidence.</td>
</tr>
<tr>
<td>• I’m not sure if this is the last question, but I’m very interested in the results.</td>
</tr>
</tbody>
</table>
## Appendix G

<table>
<thead>
<tr>
<th>S&amp;B Belief Dimensions</th>
<th>Orientations (for each dimension, ordered, top to bottom, according to an emphasis on knowledge reproduction to knowledge construction)</th>
<th>Order as it Appears in the S&amp;B Study</th>
<th>Order as it Appears in the Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Nature &amp; structure of knowledge</td>
<td>Academic views knowledge to be assessed as external to students and as a collection of important bits (definitions, concepts, techniques, methods, theories).</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Academic views knowledge to be assessed as external to students and as a coherent body of knowledge structured by experts in the field.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Academic views knowledge to be assessed as what has been internalised, reorganised, and reconstituted in the process of learning.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11. Degree of integration of knowledge</td>
<td>Academic believes that assessment should draw on information presented in a single lecture, tutorial, practical session or chapter.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Academic believes that assessment should require students to draw on information presented in many sources, but within their subject.</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Academic believes that assessment should require students to integrate information from many sources, and/or from more than one subject, and their own experience.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>12. Degree of transformation of knowledge</td>
<td>Academic believes that assessments should determine whether students can reproduce what they have been provided in lectures or textbooks, and/or practised in tutorials or practical classes.</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Academic believes that assessments should require the application of well known techniques, methods, laws, principles, or explanations to unseen standard problems.</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Academic believes that assessments should require students to apply their own understanding of concepts, principles, laws, theories to unseen, open-ended problems.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>13. Differences between good &amp; poor answers</td>
<td>Academic believes that the difference lies in the quantity of information correctly recalled.</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Academic believes that the difference lies in the accuracy and relevance of what is recalled.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Academic believes that good answers are purposeful and justify the information used, whereas poor answers do neither of these things.</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14. Role of assessment in teaching &amp; learning</td>
<td>Academic believes that students have to be forced to study, and assessment is believed to be the best tool to achieve this.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Academic believes that assessment forces students to study, and that marks give them an indication of the progress made and reward their efforts.</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Academic believes assessment to be an integral part of teaching and learning, a means of helping students learn.</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>15. Use of feedback gained from assessment</td>
<td>Academic believes that feedback from student performance should be used to alter his/her teaching.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Academic believes that feedback from student performance should be used to change the academic’s or students’ actions.</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Academic believes that feedback from student performance should be used to monitor students’ learning and to help them improve.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Academic believes that feedback from student performance should be used to challenge students’ existing ideas and understandings.</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix H

Participants’ selected course context for survey questions if other than calculus or statistics is selected:

1. Basic Algebra
2. Beginning Algebra (developmental level; 3 participants)
3. Using Mathematica
4. Classical Algebra
5. College Algebra (4 participants)
6. College Mathematics I
7. Contemporary Mathematics
8. Intermediate Algebra (5 participants)
9. Math for pre-service elementary teachers
10. Survey of Mathematics
11. Trigonometry
## Appendix I

Course representativeness 2006 vs. 2010 SUNY data

<table>
<thead>
<tr>
<th>Course Category</th>
<th>2006 SUNY Courses (60 Total)</th>
<th>2010 SUNY Courses (37 Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculus (Pre-Calculus, I, II, III)</td>
<td>13.3% (8)</td>
<td>32.4% (12)</td>
</tr>
<tr>
<td>Statistics</td>
<td>21.7% (13)</td>
<td>29.7% (11)</td>
</tr>
<tr>
<td>Developmental (Pre-Algebra, Beginning Algebra, Intermediate Algebra)</td>
<td>20% (12)</td>
<td>16.2% (6)</td>
</tr>
<tr>
<td>Algebra (Algebra, College Algebra, College Algebra &amp; Trigonometry)</td>
<td>16.7% (10)</td>
<td>10.8% (4)</td>
</tr>
<tr>
<td>Liberal Arts Mathematics (Contemporary Mathematics, Survey of Mathematics...)</td>
<td>20% (12)</td>
<td>5.4% (2)</td>
</tr>
<tr>
<td>Miscellaneous (Financial mathematics, Technical Mathematics, History of Mathematics, Numerical Methods, College mathematics I)</td>
<td>8.3% (5)</td>
<td>2.7% (1)</td>
</tr>
</tbody>
</table>
## Appendix J

### Question-by-question breakdown of ATI results

<table>
<thead>
<tr>
<th>ATI Questions</th>
<th>Measure</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(1-only rarely true, 2-sometimes true, 3-true about half the time, 4-frequently true, 5-almost always true)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scale: Conceptual change/student-focused (CCSF) approach</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subscale: Conceptual change intention items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel that the assessment in this subject should be an opportunity for students to reveal their changed conceptual understanding of the subject</td>
<td>3.9</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>I encourage students to restructure their existing knowledge in terms of the new way of thinking about the subject that they will develop</td>
<td>3.5</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>I feel that it is better for students in this subject to generate their own notes rather than always copy mine</td>
<td>3.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>I feel a lot of teaching time in this subject should be used to question students’ ideas</td>
<td>2.6</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td><strong>Subscale: Student-focused strategy items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my interactions with students in this subject I try to develop a conversation with them about the topics we are studying</td>
<td>3.7</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>I set aside some teaching time so that the students can discuss, among themselves, the difficulties that they encounter studying this subject</td>
<td>3.3</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>In teaching sessions for this subject, I use difficult or undefined examples to provoke debate</td>
<td>2.1</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>I make available opportunities for students in this subject to discuss their changing understanding of the subject</td>
<td>3.5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td><strong>Scale: Information transmission/teacher-focused (ITTF) Item no.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subscale: Information transmission intention items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel it is important that this subject should be completely described in terms of specific objectives relating to what students have to know for formal assessment items</td>
<td>3.8</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>I feel it is important to present a lot of facts to students so that they know what they have to learn for this subject</td>
<td>3.2</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>I think an important reason for running teaching sessions in this subject is to give students a good set of notes</td>
<td>2.7</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>I feel that I should know the answers to any questions that students may put to me during this subject</td>
<td>3.9</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td><strong>Subscale: Teacher-focused strategy items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I design my teaching in this subject with the assumption that most of the students have very little useful knowledge of the topics to be covered</td>
<td>3.6</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>In this subject I concentrate in covering the information that might be available from a good textbook</td>
<td>4</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>I structure this subject to help students to pass the formal assessment items</td>
<td>3.6</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>When I give this subject, I only provide the students with the information they will need to pass the formal assessments</td>
<td>2</td>
<td>1.1</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix K

<table>
<thead>
<tr>
<th>S&amp;B Question (number)</th>
<th>Original Uncorrected Comments</th>
</tr>
</thead>
</table>
| 10. Nature & structure of knowledge (7) | Although mathematical knowledge is certainly external to students and largely due to the efforts of experts (having being built by the work of countless minds throughout the centuries), the goal of the teaching process should always be to make the students internalize and reconstruct (limited parts of) it, since only in that way the concepts and techniques can really be made into useful knowledge.  
I USE PIAGET'S MODEL OF REFLECTIVE ABSTRACTION.  
Most assessment is done through online homework and paper-and-pencil tests/exams; but some assessment comes from their ability to explain concepts to others and to work with other students in the class.  
For exams, my view is closest to the option I selected—collection of definitions, etc. However, I require each online student to complete a project whose goal is internalising, reorganising and reconstituting the information.  
Students are assessed by exams given in person, as well as assignments given on “MYMATHLAB” (for this course)  
all of the above to some extent |
| 11. Degree of integration of knowledge (4) | Applications come from the book.  
This answer refers to exams... Their project requires information from their own experience as well. usually cover about 3 chapters per assessment (4 tests total); + final exam covers entire course  
The bulk of practices and quizzes assess understanding of the current section of material. So, a "typical" assessment would be one of these section assessments. However, there are also more in-depth discussions and exams that integrate multiple areas of content that would require students to draw information from multiple sources. |
12. Degree of transformation of knowledge (10)

I wanted to answer more than one option. Both "open-ended" problems and "unseen standard problems" depending on which particular assessment we are considering.

This method does not agree with my learning philosophy but it is what is provided by the publisher of the course.

I'm almost embarrassed to admit that much, but not all, of my assessment follows that indicated in the selected item above. Using the software product that we do, it's difficult for me to really know if students understand what they are doing, or just mimicking what they have viewed in the videos or the sample problems on the software. The level of conceptual teaching and assessment is much lower in my online class than in my face to face class, or at least I'm less able to determine the degree to which students really understand the material.

The students use Minitab and from the material draw conclusions. There are multiple choice, open ended, essays, critical thinking questions through out the course along with discussions.

Again, except in the project, which is unseen and open-ended.

The above describes the bulk of my assessment - but I do also include assignments in which students are asked to relate problems they select from the text to their own career interests. Also I have required online discussions that involve more open ended questions.

also use graded homework (i.e, C)

students must sue methods, etc appropriate to the level of the course and that are demonstrated in the extensive on-line notes

At this level, students are primarily doing "skill-and-drill" and only rarely do open-ended problems, or problems unlike those they haven't seen in the past.

I use a lot of open-ended questions to test for understanding too...but certainly not a majority of the questions are of this nature. The is no better way to test understanding.

---

13. Differences between good & poor answers (0)

NO COMMENTS

14. Role of assessment in teaching & learning (9)

not so clear cut. among other things, all three options are applicable

Although the answer above reflects what I believe about assessment, the course I teach online is very structured and forces students to view the assessments more as the "purpose" for their studies...with their grades/marks on each assessment weighing heavily towards their final course grade.

I'd like to be in the world where the second option applies, but day-to-day I find students left to their own devices will do as little as possible.

In an online course, assessment also helps keep students from getting behind.

the second answer is also impt

"Assessment" as I understand it is an effort to find out if the things we are doing as instructors are effective in producing learning in students. Clearly this question is referring to something else. Assessment measures what is going on; it does not force students to do anything.

At this level, I give students many, many practices and self-quizzes in order to encourage them to regularly get involved in learning the material. These practices and self-quizzes are worth only a small portion of the grade. But these practices and self-quizzes make up the bulk of the learning done in the course. Prequiz-practice/questions/reading/videos-then post quiz cycle in order to master the skills.

Assessment is the external incentive to help students do what is necessary to learn the material being presented, but in the long run the student needs to want to learn the material whether it is for external or internal reasons.

My assessment is based solely on problem sets students submit, and Mathematica is a program which can only be mastered by doing repeated examples before doing a problem
| 15. Use of feedback gained from assessment (10) | to some extent, all of the above  
I wish I could have checked the last button because philosophically, this is what I believe.  
Additionally, I do make changes to the course if students are not understanding the concepts. I make changes to my assessment tools constantly.  
Huh. I want to chose 1, 2 and, well, not I’m dithering on 3 or 4 vs 3 and 4. All of these have their strengths.  
I do all of the above  
Although I use all of these at times.  
I actually do all 4.  
All or parts of each of these answers  
also B  
This is a hard question to answer. All of them are pretty much the same. |
Appendix L

Note: Questionnaire was not seen by participants.

**Study II Interview Questions**

Participant Name: ___________________________ Date: ___________________
Start time: ___________ End time: ___________

F2F/FO Course Name:
____________________________________________________________________

FO Course Context (circle): Calculus Statistics Other (specify): ________

**F2F Instruments:**

Instr 1: ___________ @ ___% rep. OR const. Instr 4: ___________ @ ___% rep. OR const.
Instr 2: ___________ @ ___% rep. OR const. Instr 5: ___________ @ ___% rep. OR const.
Instr 3: ___________ @ ___% rep. OR const. Instr 6: ___________ @ ___% rep. OR const.

**FO Instruments:**

Instr 1: ___________ @ ___% rep. OR const. Instr 4: ___________ @ ___% rep. OR const.
Instr 2: ___________ @ ___% rep. OR const. Instr 5: ___________ @ ___% rep. OR const.
Instr 3: ___________ @ ___% rep. OR const. Instr 6: ___________ @ ___% rep. OR const.

**Introduction:** Thank you for agreeing to discuss your assessment practices with me.

FYI: There are a maximum of seven main questions with the total number I ask and the time the interview takes dependent on your particular assessment practice.

I will be recording the interview but please keep in mind, as I wrote in your email, the conversation.
will be treated as confidential, used anonymously solely for my study and you are welcome to stop the interview at any time.

Any questions before we begin?

Don’t forget... questions refer to your particular course context

Questions on the Use of Discussion

1. Use of discussions as an actual part of your assessment scheme (i.e. weighted)
   a. In your F2F teaching, do you have ‘discussion’ as a weighted part of your overall course grade? Yes or No (circle)
   b. In your FO teaching, do you have ‘discussion’ as a weighted part of your overall course grade? Yes or No (circle) If so, what do you hope to accomplish through the use of discussion?
   c. What is different about how you encourage your students to interact F2F vs. FO? If different, are there any particular reasons you can share?

Questions on the Use of Quizzes

2. Use of quizzes
   a. Do you use quizzes as part of your F2F assessment practice? Yes or No (circle)
   b. Do you use quizzes as part of your FO assessment practice? Yes or No (circle)
   c. If you answered ‘yes’ to ‘a’ and ‘b’: Is there any difference about what you hope to accomplish through the use of quizzes as part of your F2F vs. FO assessment practice? If so, are there any particular reasons you can share?
   d. If you answered ‘yes’ to ‘a’ or ‘b’ but not both: Are there particular reasons you can share why one context but not the other? What do you hope to accomplish by using quizzes as part of your assessment practice in that context?

Questions on the Use of Proctoring

3. Proctoring in the FO context:
   a. Do you use proctoring in this course when you teach F2F? Yes or No (circle)
b. Do you use proctoring in this course when you teach FO? Yes or No (circle)

c. Do you prefer proctoring? Why?

d. If answers to ‘a’ and ‘b’ are different: Which assessment instrument(s) do you proctor F2F but not FO (or vice versa)?

e. If answers to ‘a’ and ‘b’ are different: Can you share any particular reasons why you use proctor in one context and not the other?

Questions on the Use of Feedback

4. Kind of Feedback

a. Is the kind (e.g. correct/incorrect vs. full solution vs. hints/comments) of feedback you provide F2F different from the kind of feedback you provide FO? Yes or No (circle) If so, are there any particular reasons you can share?

b. What is the difference between the kind of feedback you provide F2F vs. FO? For which assessment instrument(s) is it different and how?

5. Feedback and Learning

a. Is what you are trying to accomplish with your F2F feedback different from what you are try to accomplish with your FO feedback? Yes or No (circle) If so, are there any particular reasons you can share?

b. What is the difference about what you are trying to accomplish with your F2F vs. your FO feedback? For which assessment instrument(s) is it different and how?

c. In general, what kind(s) of assessment feedback do you consider most effective in helping students understand mathematics?

6. Timing of feedback

a. Do you find that you are quicker or slower giving feedback in your FO vs. your F2F courses? Are there any particular reasons you can share?

b. Do you think immediate feedback (e.g. CAA) helps students understand mathematics? Yes or No (circle) Are there any particular reasons you can share?
Questions on Knowledge Construction

My research uses a study framework that considers how instructors tend to approach their assessment practice. *This framework considers that instructors tend to orient their assessment somewhere between an emphasis on reproducing important math facts, procedures and skills to an emphasis on the ability to purposefully use mathematical knowledge to address open-ended problems not previously encountered.*

The remaining question asks you to think about what you tend to emphasize.

7. Knowledge construction

   a. For each of your F2F assessment instruments, where would you say you tend to orient your approach? *Use table at the beginning of the questionnaire.*

   b. For each of your FO assessment instruments, where would you say you tend to orient your approach? *Use table at the beginning of the questionnaire.*

   c. If answers to ‘a’ and ‘b’ are different, are there particular reasons?

*Background Info: As there is often a mixture of questions, please consider the weight (or percent of the overall grade) rather than number of questions.*

For example, if more than half of the grade for your chapter tests tends to be questions that emphasize reproducing important math facts, procedures and skills, then you would say, for this assessment, that you tend to emphasize that side.

8. Do you think the role of assessment in F2F mathematics courses is different than the role of assessment in FO mathematics courses?

9. (If Time) Math Context: For students to understand mathematics at a deep level, what do you consider to be good and bad regarding FO mathematics assessment?

Thank you!
Appendix M

Interview participants’ assessment schemes.


<table>
<thead>
<tr>
<th>F2F (Introductory Statistics)</th>
<th>P1</th>
<th>FO (Introductory Statistics)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td><strong>Use of CAA</strong></td>
<td><strong>Weighting (%)</strong></td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>57</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>29</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3KR</td>
<td>3No-CAA</td>
<td>3</td>
</tr>
<tr>
<td>KR-100%</td>
<td>CAA-0%</td>
<td>100%</td>
</tr>
<tr>
<td>mid-0%</td>
<td>No CAA-100%</td>
<td>No</td>
</tr>
</tbody>
</table>

*For optional bonus grades. Not counted in analysis.

<table>
<thead>
<tr>
<th>F2F (Introductory Statistics)</th>
<th>P2</th>
<th>FO (Introductory Statistics)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td><strong>Use of CAA</strong></td>
<td><strong>Weighting (%)</strong></td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>30</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>KR</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>kr*</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3KR-1mid-1KC</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>KR-80%</td>
<td>CAA-20%</td>
<td>100%</td>
</tr>
<tr>
<td>mid-5%</td>
<td>No CAA-80%</td>
<td>No</td>
</tr>
<tr>
<td>KC-15%</td>
<td>No CAA-80%</td>
<td>No</td>
</tr>
</tbody>
</table>

*Though viewed as less oriented to knowledge reproduction than ‘paper-based’ homework.
**Inferred from the participants’ account of how discussion is used.

<table>
<thead>
<tr>
<th>F2F (Calculus)</th>
<th>P3*</th>
<th>FO (Mathematica for Calculus)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td><strong>Use of CAA</strong></td>
<td><strong>Weighting (%)</strong></td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>80</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>KR</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Instrument Count</td>
</tr>
<tr>
<td>KR-10%</td>
<td>CAA-10%</td>
<td>100%</td>
</tr>
<tr>
<td>mid-80%</td>
<td>No CAA-90%</td>
<td>100%</td>
</tr>
</tbody>
</table>

*The only participant not referring to the same course in both contexts.
<table>
<thead>
<tr>
<th>Approach (KR/mid/KC)</th>
<th>Use of CAA</th>
<th>Weighting (%)</th>
<th>P4</th>
<th>Weighting (%)</th>
<th>Use of CAA</th>
<th>Approach (KR/mid/KC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mid</td>
<td>No</td>
<td>25</td>
<td>Final Exam</td>
<td>20</td>
<td>Yes</td>
<td>KR</td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>50</td>
<td>Exams</td>
<td>50</td>
<td>Yes</td>
<td>KR</td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>10</td>
<td>Quizzes</td>
<td>10</td>
<td>Yes</td>
<td>KR</td>
</tr>
<tr>
<td>KR</td>
<td>Yes</td>
<td>8</td>
<td>Homework</td>
<td>7.5</td>
<td>Yes</td>
<td>KR</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>4</td>
<td>Group Project</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>3</td>
<td>Lab Activity</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Vista</td>
<td>8</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Discussion</td>
<td>4.5</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KR-8%</strong></td>
<td><strong>CAA-8%</strong></td>
<td>100%</td>
<td><strong>Total</strong></td>
<td>100%</td>
<td></td>
<td><strong>KA-87.5%</strong></td>
</tr>
<tr>
<td><strong>mid-85%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>mid-65%</strong></td>
</tr>
<tr>
<td><strong>KC-7%</strong></td>
<td>No CAA-92%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td><strong>No CAA-12.5%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach (KR/mid/KC)</th>
<th>Use of CAA</th>
<th>Weighting (%)</th>
<th>P5</th>
<th>Weighting (%)</th>
<th>Use of CAA</th>
<th>Approach (KR/mid/KC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kc*</td>
<td>No</td>
<td>40</td>
<td>Exams (Invigilated)</td>
<td>50</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>40</td>
<td>Exams (Take Home)</td>
<td>35</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>20</td>
<td>Quizzes</td>
<td>10</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td>mid</td>
<td>No</td>
<td>0</td>
<td>Homework</td>
<td>5</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>Instrument Count</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KR-0%</strong></td>
<td><strong>CAA-0%</strong></td>
<td>100%</td>
<td><strong>Total</strong></td>
<td>100%</td>
<td></td>
<td><strong>CAA-0%</strong></td>
</tr>
<tr>
<td><strong>mid-20%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>mid-65%</strong></td>
</tr>
<tr>
<td><strong>KC-80%</strong></td>
<td>No CAA-100%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td><strong>No CAA-100%</strong></td>
</tr>
</tbody>
</table>

* Though viewed as less oriented to knowledge construction as the take home exams.

<table>
<thead>
<tr>
<th>Approach (KR/mid/KC)</th>
<th>Use of CAA</th>
<th>Weighting (%)</th>
<th>P6</th>
<th>Weighting (%)</th>
<th>Use of CAA</th>
<th>Approach (KR/mid/KC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mid</td>
<td>No</td>
<td>50</td>
<td>Exams</td>
<td>46.2</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>16.6</td>
<td>Projects</td>
<td>7.7</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>11.1</td>
<td>Homework</td>
<td>8.2</td>
<td>No</td>
<td>KR</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>5.6</td>
<td>Journal</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>5.6</td>
<td>Discussion (Online)</td>
<td>5.5</td>
<td>No</td>
<td>mid</td>
</tr>
<tr>
<td>KR</td>
<td>No</td>
<td>5.6</td>
<td>Quizzes</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KC</td>
<td>No</td>
<td>5.6</td>
<td>In Class Participation</td>
<td>5.5</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Preliminary Assignment</td>
<td>1</td>
<td>No*</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Written Assignments</td>
<td>7.3</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Exploratory Assignments</td>
<td>9.2</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Virtual Class</td>
<td>9.2</td>
<td>No</td>
<td>KC</td>
</tr>
<tr>
<td>2KR-1mid-3KC</td>
<td>CAA-0</td>
<td>7</td>
<td>Instrument Count</td>
<td>8</td>
<td>CAA-0</td>
<td>KR-8.2%</td>
</tr>
<tr>
<td>KR-16.7%</td>
<td>CAA-0</td>
<td>100.1**</td>
<td><strong>Total</strong></td>
<td>99.8**</td>
<td>CAA-0</td>
<td><strong>mid-51.7%</strong></td>
</tr>
</tbody>
</table>

* Although, ‘one part is CAA’.

** Rounding errors due to percentage conversion from the original point-based system.
Overall summary of interview participants’ assessment schemes.

<table>
<thead>
<tr>
<th>Approach to Assessment*</th>
<th>F2F Course</th>
<th>FO Course</th>
<th>Least KR and Most KC</th>
<th>Difference (KR/m/KC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KR mid KC</td>
<td>KR mid KC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>100% CAA</td>
<td>0 0</td>
<td>100% CAA</td>
<td>tie</td>
</tr>
<tr>
<td>P2</td>
<td>80 (20% CAA)</td>
<td>5 15</td>
<td>80 (20% CAA)</td>
<td>virtual tie</td>
</tr>
<tr>
<td>P4</td>
<td>8% CAA 85 7</td>
<td>87.5 CAA 4.5 8</td>
<td>F2F</td>
<td>-79.5/+80.5/+1</td>
</tr>
<tr>
<td>P5</td>
<td>0 20 80</td>
<td>0 65 35</td>
<td>F2F</td>
<td>tie/-45/+45</td>
</tr>
<tr>
<td>P6</td>
<td>16.7 50 33.4</td>
<td>8.2 51.7 38.9</td>
<td>virtual tie</td>
<td>-8.5/+1.7/+5.5</td>
</tr>
<tr>
<td>Average</td>
<td>40.94 32 27.08</td>
<td>55.14 24.24 20.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Approach to Assessment:
- F2F: Face-to-Face
- FO: Online Format

**Note:** The difference in KR and KC values indicates the perceived differences in knowledge requirements and commands for each approach.