

**An investigation of assessment and feedback practices in fully asynchronous
online undergraduate mathematics courses**

Sven Trenholm

University of South Australia, AU

Lara Alcock and Carol Robinson

Loughborough University, UK

Correspondence concerning this article should be addressed to:

Sven Trenholm

School of Education

University of South Australia

Mawson Lakes Campus

Garth Boomer Building

Office G2-27

Adelaide, SA 5095

Email: sven.trenholm@unisa.edu.au

Tel.: +61 (08) 830 26358

Fax: +61 (08) 830 26688

Abstract

Research suggests it is difficult to learn mathematics in the fully asynchronous online (FAO) instructional modality, yet little is known about associated teaching and assessment practices. In this study we investigate FAO mathematics assessment and feedback practices in particular consideration of both claims and findings that these practices have a powerful influence on learning.

A survey questionnaire was constructed and completed by 70 FAO undergraduate mathematics instructors, mostly from the US, who were each asked to detail their assessment and feedback practices in a single FAO mathematics course. Alongside these questions, participants also answered the 16-item version of the Approaches to Teaching Inventory. A novel feedback framework was also created and used to examine how feedback practices may be related to participants' approaches to teaching.

Results show that assessment and feedback practices are varied and complex: In particular, we found there was not a simple emphasis on summative assessment instruments, nor a concomitant expectation these would always be invigilated. Though richer assessment feedback appears to be emphasized, evidence suggests this feedback may not be primarily directed at advancing student learning. Moreover, we found evidence of a reliance on computer-human interactions (e.g. via CAA systems) and further evidence of a decline in human interactions, suggesting a dynamic that is both consistent with current online learning theory and claims FAO mathematics courses are becoming commodified. Several avenues for further research are suggested.

Keywords: online learning, tertiary mathematics, feedback, assessment, approaches to teaching

Classification code: 97B40 97U50

1. Background

Faced with an increased demand for higher education, institutions and government policy makers are turning to online courses as an effective and efficient means of delivering instruction.[1,2] Reflecting this interest, growth in these courses is considered to be “exploding” [3] with “online” enrollments at US degree-granting tertiary institutions reportedly growing at an average annual rate of about 14% for the five years from 2006 to 2010.[4] Consistent with these trends, both the demand for [5] and development of online *mathematics* courses is increasing, [6,7] and substantial investments of both time and money are being directed at related course development efforts.[8]

The most prevalent method of online instruction is fully asynchronous online (hereafter “FAO”) (see, for example, [9]) in which all interactivity is mediated by the Internet, with instructors and students separated in both space and time¹. In mathematics, the latest US Conference Board of Mathematical Sciences (CBMS) survey reports that 35% of four-year mathematics departments offered distance-learning courses of which 72% are “completely online” and 88% of two-year colleges offered distance-learning courses of which 73% are “completely online”. [10; see Table SP.10]

However, mathematics has been singled out as a challenging discipline to teach in this modality.[7,8,11-15] In particular, most meta-analytic research, typically involving some form of FAO vs. traditional face-to-face (hereafter “F2F”) course comparison of student achievement measures, presents general FAO instruction in a fairly positive light but, upon closer inspection, appears to question its viability for mathematics.(see, for example, [16-18])

With the focus of this research almost entirely on output elements such as student grades, it is unclear how input elements such as instructors’ assessment and feedback practices factor in or compare. In this study we consider these practices in FAO mathematics courses. We do so in particular consideration of reflective claims that course assessment practices have a powerful influence in directing student learning [19-22], and both reflective claims [23,24] and empirical research findings [25] that feedback has a powerful influence in advancing that learning. In short, we argue that FAO assessment and feedback practices act as vital instructional inputs and

¹ There are some exceptions: Remote students may, for example, phone to ask for help. On campus students may, for example, use “office hours” to ask for help.

related investigations will help provide some needed insight into the nature of current FAO mathematics instruction.

1.1 *The practice of feedback*

The general literature considers feedback to be a distinguishing characteristic of assessment instruments [26] which is critical to advancing student learning.[25] We use Ramprasad's [27] definition of 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). For the purposes of this study, this is information given by the instructor to the student *in the context of assessment*, and includes automated computer-assisted assessment (hereafter "CAA") feedback. We define formative assessment as linked with *feedback* practices that are concerned with closing any gap in learning (i.e. a developmental process; [28]) and summative assessment as linked with the idea of *feedout* where the gap in student learning is simply measured with no specific intention to address any learning needs (i.e. a judgement of the product; [29]).

Despite its stated potential, the possible effects of feedback on learning should not be oversimplified as wide variability has been reported, with some feedback found to produce no or debilitating effects.[25,30] Furthermore, differences in effect have been found, for example, to be based on how frequently feedback is provided (e.g. number of times feedback is given over the course of attempting one question; [25]), how specific it is (e.g. addressing "correctness of the minutiae of tasks"; [25,p. 91;31]), the nature of the subject matter being studied [30] and the characteristics of the student receiving the feedback.[32] However, overall, two characteristics of feedback are considered to have the most influence on learning: timing and kind.

1.1.1 *Feedback timing*

Despite claims that the provision of "rapid" feedback "promote[s] active, deep learning",[33,p.11] the effect of feedback timing, as indicated in the general literature, is not so simple. Typical studies focus upon the effects of immediate versus delayed feedback and findings suggest that timing effects depend upon how difficult an individual student finds a particular question; for example, the greater the need for students to process the material to gain understanding the more they may benefit from delayed feedback.[30] Moreover, at least one meta-study concludes that the benefit of immediate feedback may be limited to lower-level "procedural skills" and providing motivation.[34,p.165] Such findings suggest a greater degree

of complexity than simply stating that immediate (or delayed) feedback is beneficial to student learning.

In mathematics, there has been limited related research. In one study of whole class instruction (n=20 classes), Tobin,[35] for example, found “higher mathematics achievement and improvements in the quality of teacher and student discourse” to be associated with an average feedback timing of between 3 and 5 seconds (p.191). In another study based on a comparison of problem solving using a computer software package versus paper (n=64 children), Simmons and Cope [36] 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. However, in relation to CAA feedback, others have claimed that immediate feedback is linked with “gaining a thorough understanding” (p.56) of mathematics [37] or found that immediate CAA feedback enables increased grade performance,[38] both of which are consistent with prior research *if* related performance effects refer to lower-level or procedural understanding.[34] In sum, the balance of evidence would appear consistent with general findings which suggest immediate feedback would be beneficial to procedural learning whereas delayed feedback would be beneficial to conceptual learning. This will be further discussed in the coming section on CAA feedback.

1.1.2 Feedback kind

The quality of feedback provided by instructors has been directly linked to developing deep student understanding,[23,39] which might be considered consistent with conceptual understanding in mathematics. Qualitatively different kinds of assessment feedback may be characterized in several ways. As shown in Table 1, it could be said the first kind, considered the poorest quality assessment feedback, consists solely of a grade or mark. Such feedback is widely considered [40] and found [41] to be the least beneficial to student learning. For example, it does not provide any direction to help further learning [42] and is possibly detrimental.[43] At the other end of the continuum, the kind of feedback widely considered [44] and found [25] to be most beneficial is hints and comments directed at the learning process. That is, feedback directed at the learning task and the development of student understanding.[44] The final intermediate kind of feedback consists of providing the correct answer or a full solution (whether computer- or instructor-generated). This feedback has been found to be better than the first category [30] but is not considered to be better than the third category [45 as cited

in 46]. Together, these claims and findings may be represented by a taxonomy of feedback quality relative to learning effect.

[Table 1 near here]

This taxonomy appears consistent with findings in mathematics education. 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”. [50,p.43] Additionally, in geometry, CAA feedback that provided hints was found to improve learning, as measured by pre- to post-test gains, significantly more than feedback that showed answer correctness or the full solution. [51] Finally, the only known meta-analysis on feedback practice in mathematics concluded that specific feedback on academic performance was found to consistently “enhance mathematics achievement”. [52,p.67]

1.1.3 FAO mathematics feedback practices

In the general FAO literature, the use of feedback is emphasized, [53-57] particularly as effective FAO assessment practice is often associated with the use of formative [55,57-59] and varied assessment instruments. [54,56,59-61] However, reflecting the general state of undergraduate mathematics research and the nascent state of FAO instruction, little is known about undergraduate mathematics F2F feedback practices in general and FAO practices in particular [62]: In F2F undergraduate mathematics instruction, there is an emphasis on summative assessment, [8,20,29,63-66] and assessment feedout (rather than feedback) processes are expected to dominate. In FAO mathematics instruction, of what very little is known, feedback practices appear much more complex [67] and many claim an extended purpose for the use of feedback. Chiefly, it is seen as a vital means of directing student learning and mediating for the nature of the FAO course context where students are separated from each other and the instructor in both space and time. [68,69] FAO feedback might, for example, be the only or primary contact students have with their instructor [70]; linked with this assertion, good feedback is considered as a means to “stimulate” [60,p.2341] and “maintain” interactions [71 as cited in 72,p.81]. As Semião [73] argues: it serves to “keep students engaged with the process...of learning” (p.439). On balance, such feedback appears to serve two distinct, yet complimentary purposes: that of helping to advance as well as direct student learning.

Finally, perhaps more importantly, the ability to time feedback to advance learning, such as was discussed earlier, is almost impossible in an asynchronous environment.[74] Instead, generalized FAO findings link feedback that is, for example, “prompt”, “frequent” or “immediate” to “learning effectiveness”². [75,p.24] In short, immediate FAO feedback is generally not seen as a possible threat to higher quality learning.

1.2 *FAO undergraduate mathematics instructors’ assessment practices – instruments, weighting and delivery*

In order to gain a broader perspective, we turn to a review of the literature on assessment practices in FAO mathematics courses in which the research is both complicated and limited. There are problems with assessment terminology and delivery: The role and method of delivery of a particular assessment instrument may vary from instructor to instructor and instrument to instrument³. For example, one instructor’s “quiz” may be another instructor’s “test”, one instructor’s “quiz” may be paper-based while another instructor’s “quiz” may be delivered via CAA and, if both use CAA, the nature of the questions and feedback may vary depending on, for example, the computer platform. Furthermore, in FAO courses, instruments typically assumed to be delivered under controlled conditions (e.g. a final exam with human supervision allowing limited time) are often delivered under much less controlled conditions (e.g. a “take-home” final exam with no human supervision allowing considerably more time).[67, 76] With these limitations in mind, the following section provides a brief review of research in this area with a particular focus on the nature of assessment schemes, which we define as the combination of instruments used together with their associated weighting. These schemes are where we assume the power of assessment (to drive and direct learning [19-22, 77]) is operationalized.[78]

As previously discussed, FAO mathematics assessment practice is considered to be different from [61,70, 79, 80] and more complex than F2F practice.[70] However, though there is a growing body of literature on the use of different formative assessment instruments in FAO mathematics courses (for example: peer assessment [81]; group work or projects [70]; discussion [82]; journal writing [83]), actual course assessment practices appear to mirror F2F practices where summative instruments prevail. For example, Galante,[84] in her survey of 37

² Learning effectiveness in this study is considered as a measure of how institutions FO learning is equivalent to or better than F2F learning.

³ 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, if retakes are available or if corrections can be submitted for extra credit.

(of 472 or 9% response rate) FAO mathematics instructors from around the US, found “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). Similarly, Smith, Heindel, and Torres-Ayala,[85] in their log file analysis of 406 online courses at one US public university, found that hard-pure FAO courses (e.g. mathematics) use tests and question pools more than other disciplines. However, reflecting some of the complexity, Trenholm,[67] in his survey of 47 (of 122 or 39% response rate) FAO mathematics instructors in one US award-winning virtual learning environment, found those using invigilation (i.e., proctored testing) generally placed an emphasis – in terms of usage and weighting – on summative instruments while those not using invigilation did not. Whether instructors had freely chosen not to invigilate, or it was an administrative mandate (i.e. institutions looking to uphold the “‘anytime, anywhere’ creed of online instruction” [76, p.291]), he argued that instructors who did not use invigilation were replacing traditional summative assessment instruments with formative-style assessment instruments and questioned whether this shift was related to improving student learning outcomes.

1.2.1 Use of CAA

There are claims but little empirical evidence that the use of CAA is high among FAO mathematics instructors.[8,86] CAA instruments, such as online quizzes, are often characterized by the provision of quick or immediate feedback, by frequent administration (e.g. weekly) and by the use of randomized questions that provide multiple attempts to answer (termed an “attempt-feedback-reattempt system” [38,p.132].[87-89] To the extent that CAA instruments are associated with immediate feedback and considered to fulfill a primarily formative role,[89] they appear aligned with current general FAO assessment practices as previously discussed.

However, there appear to be competing claims and findings concerning the potential of CAA instruments to affect learning. Some, for example, consider the use of CAA as beneficial to mathematics instruction,[37, 90] leading to improved performance.[38,88] Others, in seeming contrast, have found that it can only address lower level or procedural learning.[91,92] On balance, these findings appear to suggest any benefit might be limited to lower level or procedural learning. Indeed, given the association of immediate feedback to CAA systems, this is consistent with wider research linking the benefit of immediate feedback to procedural learning.[34]

There are, moreover, reasons to suggest that current CAA systems and feedback may also be detrimental to student learning. With evidence of learning generally limited to only a single answer, and not a progression of steps needed to solve a problem, it is difficult for CAA systems to accurately judge student thinking.[93,94] The provision of repetitive attempts has also been linked to the potential reinforcement of “incorrect interpretations” [95,p.28] and trial-and-error [36] or “automatic strateg[ies with] no underlying mathematical understanding”.[96,p.243] In short, some students engaged with this interactivity could be arriving at the right answer but wrong understanding.

1.2.2 Use of invigilation

Finally, of the limited number of studies on the use of invigilation in FAO mathematics courses, the strongest evidence suggests it is fairly common not to use any invigilation, with use or non-use found to be associated, as previously discussed, with a de-emphasis on summative-style assessment instruments.[67] This may be expected given the anonymity afforded by the FAO context makes it relatively easy to cheat, particularly in math or fact-based courses.[67,73,76,79,97,98] Still, among FAO mathematics instructors, some argue for [99] and others against [100] the need for test supervision. The contentious nature of this issue was laid bare when Yates and Beaudries’ [100] study prompted a response article [101] outlining five problem areas which effectively nullified the initial study findings. Nonetheless, there remains no conclusive evidence that the use of invigilation is unnecessary, while concerns remain regarding the validity of non-invigilated summative assessments.

1.3 Summary

The use of feedback – kind and timing – have been found to have a significant influence on advancing student learning and, as such, may be considered important defining characteristics of assessment instruments. However FAO mathematics assessment and feedback practices appear more complex than F2F practices: It is claimed FAO assessment feedback practices are used for directing and not simply advancing student learning, that the use of CAA and associated computer-generated feedback is strongly relied upon, and that more formative-style assessment instruments are used when courses do not use any invigilation. Together, these issues may suggest ongoing efforts to mediate for the nature of the FAO context; but what is actually happening, to assessment and feedback practices in particular, remains unclear. We thus pose

two broad questions:

1. What assessment instruments are FAO mathematics instructors currently using to assess their students? How are these weighted? Which are invigilated?
2. How are FAO mathematics instructors using feedback with these instruments? What kind of feedback? How is it timed?

2. Methodology

Survey questions were developed to investigate FAO undergraduate mathematics instructors' assessment and feedback practices. These questions were part of a larger mixed methods (quantitative followed by qualitative) study [102] where findings from a survey completed by 70 FAO undergraduate mathematics instructors were explored using semi-structured interviews with six purposefully selected survey participants.[103] This paper represents the quantitative research component of the larger study while the latter represents the qualitative. The full survey was structured in four parts – participant demographics, approaches to assessment, assessment specifics (which included assessment instrument weightings and associated feedback practices) followed by approaches to teaching. The present study details the results related to participants' course assessment specifics and how they are related to approaches to teaching. Following questions on participant demographics, participants were asked to select a single FAO mathematics course they taught, targeting courses at the introductory level (e.g. Introductory Statistics, Calculus), then for each identified assessment instrument, nine questions were asked with two targeting feedback practices (see Table 2).

[Table 2 near here]

As part of the larger survey and as an aid in this investigation, each participant also answered the 16-item version of the Approaches to Teaching Inventory (ATI)[104] for the same specified course context. The ATI has generally been found to be a reliable and valid psychometric instrument.[105,106] It provides two scale measures which gauge how instructors approach their teaching. The first scale measures to what degree an instructor's approach is "aimed at conceptual change" and student-focused (CCSF). The second scale measures to what degree an instructor's approach is "geared towards information transmission" and teacher-focused (ITTF).[106,p.60] Both scales provide a measure from a minimum of 8 to a maximum of 40 of how instructors are oriented to the respective scale description. To fit the FAO instructional

context one minor change was made to the wording for one question: “I set aside some teaching time” was changed to “I plan my teaching”. In the present study context, the ATI measures were used to analyse the relationship between approaches to teaching and characteristics of assessment and feedback practices “perceived by the same teacher in the same context”. [104,p.421]

2.1 Data collection

The full survey was launched using Bristol Online Surveys (<https://www.survey.bris.ac.uk>). This was done in stages beginning March 17th, 2011 with all surveys closed by May 31st, 2011. Participants were targeted using prior contacts in the undergraduate mathematics education community. This included mathematics instructors from two recognized US leaders in FAO instruction – the State University of New York (SUNY) Learning Network (SLN) and the Tennessee Board of Regents system (RODP) – researchers involved in a recent related international research compilation, [107] and members of the Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education (SIGMAA on RUME). Importantly, as shown in Table 3, most participants teach in the US public higher education context, the US context for the vast majority of US FAO instruction. [108] Finally, while online surveys were viewed as a convenient and efficient means of collecting data [109] some caution concerning generalizability is noted given the nature of online surveying [110] and issues associated with non-probability convenience sampling employed in this study.

[Table 3 near here]

Although 70 participants completed the survey, inconsistent and/or missing information caused the full data set to consist of 66 participants.

2.2 Analysis

The analysis relied upon the use of the ATI and the feedback framework. Prior to using the ATI, we simply used Cronbach’s alpha to test for the reliability of the two ATI measures. However, to undertake the analysis of feedback practice, the feedback framework presented in the background section first needed to be operationalized.

Findings and claims in the literature (see Table 1) were used to construct a novel scoring system to help analyse 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 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 4). In this respect, a single participant may acknowledge using, for example, three different assessment instruments and providing specific kinds of feedback with each of these three instruments. In addition, a single instrument may have been associated with one or a combination of different kinds of feedback, where the richest form was considered the feedback type associated with an instrument. Thus, by making use of this framework, we created a new way of viewing individual instruments and participant assessment practices.

[Table 4 near here]

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 and the average of these scores was what constituted the average feedback score for quizzes. Ranging from 0 to 2, instruments closer to 0 are considered to provide “poorer feedback” potentially leading to poorer quality learning and those closer to 2 are considered to provide “richer feedback” potentially leading to better quality learning. Finally, 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 different kinds of feedback relative to their overall assessment scheme. For example, one participant used four assessment instruments 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. This participant’s three feedback measures were: 60 for type 0, 30 for type 1 and 10 for type 2. Conceptually, we would expect participants with

high CCSF scale measures to make heavy use of type 2 feedback and those with low CCSF scale measures to make less use of type 2 feedback.[32]

While the literature formed the basis for the construction of this framework, a few validity issues are recognized. First, the framework relies primarily on claims and findings from generic de-disciplined research (as discussed earlier, there is only limited mathematics-specific research). Second, the “computer-” or “lecturer-generated hints or comments” description were open to interpretation. In particular, 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.[25] Therefore, in terms of potential effects on learning, type 2 feedback may have been operating like type 0 feedback. Similarly, if more than one type of feedback was used with a single assessment instrument, it is unclear, compared to other types, the extent to which each form of feedback was actually used. With the latter two issues, we would expect future research to specify whether the feedback is, for example, directed at students’ mathematical thinking, and that participants be further questioned with regards to how the identified feedback was used.

3. Results and discussion

The results and discussion are reported in three parts. First, we analyse the distribution of ATI measures and briefly present these findings. Second, excepting the use of feedback, descriptive statistics are detailed concerning course assessment instrumentation used (e.g. weighting, frequency, use of invigilation...). Third, descriptive statistics are provided and analyses conducted, using the ATI measures and operationalized feedback framework, concerning the kind and timing of feedback used, overall, as well as the use of feedback with CAA and invigilation. For the latter two parts, the analysis was conducted from two perspectives: assessment instrument and participant.

3.1 *ATI measures*

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. Out of a possible range of 8 to 40, participants’ CCSF and ITTF measures ranged, respectively,

from 10 to 39 and 16 to 36, with a mean of 26.0 (SD = 5.9) and 26.8 (SD = 4.5). Both measures were found not to be significantly different from a normal distribution (CCSF: $W(66) = 0.970$, $p = 0.109$; ITTF: $W(66) = 0.979$, $p = 0.314$).

Using Cronbach's alpha, the coefficients for the two ATI scales were 0.784 (CCSF) and 0.506 (ITTF). This suggested the inventory had good statistical validity for the CCSF scale.

However, the low alpha value for the ITTF scale casts a degree of doubt on the meaning of any findings related to the ITTF scale. For the remaining analysis, we therefore only make use of CCSF measures.

3.2 Use of assessment instruments

From an instrument perspective, Table 5 provides information regarding the number of participants using each of 12 assessment instruments (including one "other" category) as well as statistics on instrument weighting.

[Table 5 near here]

In terms of weighting, most FAO mathematics instructors tend to emphasize summative instruments (final exam, tests and mid-term exam), mirroring the emphasis in F2F courses.[8,20,29,63-66,111] 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, though weighting has notably high variability (SD = 26.4). Despite expectations discussed in the literature review,[86,112] the use of quizzes does not appear to be a central characteristic of FAO instruction, with little more than half (35 of 66 or 53%) the participants using quizzes as a weighted component of assessment. However, this may reflect an issue of terminology as one participant's homework may be another's quiz. There is also some evidence of "non-traditional" instruments in use, most notably the use of discussion and individual projects, though group-oriented instruments (group projects and group work activities) as well as individually-oriented instruments (journaling and portfolio work) do not appear to be used very often. Given the low participant numbers found in these latter two instrument categories, the remainder of this report focuses on the top eight identified instruments.

For the use of invigilation, our findings provide a fine-grained instrument-by-instrument view of invigilation practices in these courses, extending prior research which has simply identified

whether or not any invigilation has been used in a single course. These findings (see Table 6) indicate that only a minority of participants use invigilation with invigilated tests (11/38 or 29%), on average, weighted 20% more than those who do not use invigilation (57% compared to 36%). However, while the majority of participants (33/45⁴ or 73%) administering final exams use invigilation, some (12/45 or 27%) do not invigilate these exams and, perhaps surprisingly, there is little difference in the average weighting (34% compared to 31% respectively).

[Table 6 near here]

From an individual participant perspective, Figure 1 displays participants' assessment schemes (ordered left to right by homework and then final exam weighting from highest to lowest). This illustrates the variety of different assessment schemes in use by participants. 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, with three participants' assessment schemes composed entirely of homework. Finally, the use of multiple instruments and variation across schemes is consistent with previous findings,[67] with the number of different instruments used per participant ranging from one to seven, with a median of four. From an individual participant 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%) were found not to use any form of invigilation.[67] The use of invigilation will be further explored in the next section.

[Figure 1 near here]

3.3 *Use of feedback*

From an instrument-level perspective, Table 7 and Table 8 provide descriptive statistics regarding the nature of feedback used by participants in their assessment practice. Referring to Table 7, the type of feedback associated with each assessment instrument is summarized in two ways. The average feedback score associated with each instrument is given in column three and, for each instrument, the number of participants using each feedback type is detailed in the remaining three columns.

⁴ Number discrepancy due to missing data.

[Table 7 near here]

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 associated with final exams (0.52). Such results are perhaps to be expected because work pursued *during* a course may be expected to be associated with richer feedback whereas work administered *at the end* of a course is likely not. They also appear to support prior assumptions [67] that homework is primarily intended for formative purposes while final exams are primarily intended for summative [26].

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, we make several observations. 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 identified by participants, results suggest differing assessment purposes (i.e. formative *to* summative) for same *named* assessment instruments. However, some instruments did show tendencies to one primary purpose, evidence that appears consistent with claims of “blurring” regarding assessment purposes.[60,113,114]

Table 8 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. Both feedback frequency and timing appears to be in line with expectations. For example, homework and quizzes are assigned more frequently with quicker feedback while exams are much less frequent with slower feedback.

[Table 8 near here]

From a participant perspective, Figure 2 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. 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 0 or only type 1 feedback). On balance, participants emphasize the use of type 2 feedback with most using a combination of two or more types. Though consistent with an emphasis on formative assessment in FAO instruction,[58] findings appear to conflict with an emphasis on summative assessment instruments detailed in the previous section. However, no statistically significant correlations were found between weightings associated with the poorest and richest feedback (i.e. type 0 and 2) and the CCSF scale measures, suggesting participants' choice of feedback is not related to whether or not they are oriented to seek conceptual change with a student focus.

To further explore this, we investigated the relationship between participants' approaches to teaching and their homework feedback practices. Homework was selected because, as shown in Table 7 and 8, the average feedback score and frequency appear to reflect an assessment instrument that is widely used in the aid of developing students' mathematical understanding. We also narrowly focused on one feedback sub-category ("hints or comments [that] challenge student understanding"; see Table 2 question 9) considered most associated with developing student understanding.[25] With participants grouped according to whether or not they identified providing this feedback with their homework (n=41 providing vs. n=14 not providing), no significant difference in CCSF measures was found. That is, participants identified by the ATI as being more oriented to seek conceptual change and be student focused did not provide more of this kind of feedback than those that were not. Though limited by the sample size, this may suggest at least a couple of things: First, this feedback is not directed, at least primarily, at advancing learning. This is consistent with claims that some FAO feedback is being used to stimulate or maintain interactions, not for actual learning. This may also suggest problems with "feedback literacy" among participants;[115,p.26] that is, despite, for example, being more oriented to conceptual change and student focused, participants may not understand how to use feedback effectively. Second, there may be problems with the analysis such as, for example, limitations with respect to the feedback framework or assumptions made with respect to assessment weighting (for the former test). In support of the first suggestion, some evidence

that feedback is not directed at advancing learning was uncovered in the second phase interviews.[103] For example, when asked to identify any differences in the *kind* of feedback they used in their F2F versus their FO courses, most interview participants repeatedly identified the question with the *process* (e.g. frequency or quantity) of feedback provision – suggesting participants connect the quality of their feedback practice primarily with *how* and not, as emphasized in the literature,[25] with *what* kind of feedback is provided. This will be further discussed in the next two sections.

[Figure 2 near here]

3.3.1 Use of computer-generated feedback

Table 9 provides an instrument perspective on the use of computer-generated feedback. This was feedback identified as “computer-generated full solution” or “computer-generated hints and comments” but also included feedback identified as immediate⁵. As shown, a majority of assigned homework and quizzes used computer-generated feedback and, with the notable exception of homework, CAA and non-CAA weightings generally appear very similar.

[Table 9 near here]

From a participant perspective, the overall mean weighting of instruments identified as providing computer-generated feedback was 29.1% (SD = 31.9) with 46/66 or 70% using computer-generated feedback and 5/66 or 8% providing computer-generated feedback with all their assessment instruments. No relationship was found between participants’ assessment weighting associated with CAA and their CCSF scale measures.

With the assumption that the use of computer-generated feedback implies the use of online homework or testing software, we can compare these results with available related US statistics.[116] As shown in Table 10, and despite some limitations⁶, participants appear to be

⁵ Attesting to its association with CAA systems, 94% of all feedback in the former two categories was also identified as immediate.

⁶ There are two limitations to this comparison: First, it was gained by asking department heads, not actual instructors, to estimate the “percentage of sections” taught using “commercial or locally produced online-response homework and testing systems” (for two-year mathematics departments) or “online homework generating and grading packages” (for four-year mathematics and statistics departments). Second, the use of CAA is expected to have increased since 2005.

using CAA substantially more in FAO courses than what is happening overall in F2F courses, a claim further supported by follow up interview responses⁷. [103]

While an increased dependence on CAA systems (and concomitant use of immediate feedback [38]) makes sense given the importance of prompt feedback in FAO courses, such a reliance may also suggest these courses are addressing a lower level of learning [34,91] than F2F courses. As a question for future research, this may also suggest an additional reason why earlier tests did not find a relationship between CCSF measures and the use of feedback: participants' feedback practices may be constricted by CAA system parameters.

[Table 10 near here]

Moreover, a dependence on CAA systems is consistent with claims of commodification in FAO mathematics courses [85] as well as research that has shown problems with human interactions in FAO mathematics courses – both student-student interactions [117,118] and, across all human interactions, persistent problems with mathematical communication. [13,70] Indeed, a significant decline in weighting as well as usage of discussion as a weighed assessment instrument was found when comparing SUNY participant data from this study with data from an earlier 2006 study of SUNY courses. [67] With a similar mix of courses, 2010 discussion weighting (Mdn = 1.00) was found to be significantly less than 2006 discussion weighting (Mdn = 10.00; $U=761$, $Z= -2.5$, $p=0.012$), representing an 18.6% decrease in the number of courses using discussion as a weighted assessment instrument. This is further consistent with recent school-level Australian research which found distance teaching of mathematics is primarily a “one-on-one” experience with little use of collaboration. [119] We argue a possible reliance on CAA systems reflects efforts to present a viable alternative pathway for instruction when no synchronous (either live F2F or virtual) interaction exists.

Finally, these changes reflect a dynamic that is consistent with current online and distance education theory. [120,121] That is, current FAO mathematics instruction may be evolving such

⁷For example:

P4: *Of course...in the online course the quizzes and exams are all on the online software... and in the F2F, of course, it's paper and pencil.*

P5: *... [In FAO 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.*

that problems with human interactions (i.e. instructor-student, student-student) are leading to a greater dependence on student-technology interactions. This evolution may also be encouraged by advances in the functionality and nature of CAA systems and feedback (for example, emerging adaptive forms tailored to individual students [122]), though there still exists serious questions about the level of learning these systems can help students attain. Further research should be directed at understanding the nature of interactions in these courses, particularly in relation to the quality of learning.

3.3.2 *Use of feedback related to the use of invigilation*

Finally, we return to the use of invigilation and address claims [67] that those choosing not to use any invigilation are placing a greater emphasis on formative-style assessment instruments. Figure 3 displays participants' assessment practices by type of feedback use with the left side displaying those participants using invigilation and the right side displaying those participants not using any invigilation. As shown, the "no invigilation" group (right side) appears to make greater use of the richer type 2 feedback than those who do invigilate. Indeed, tests comparing weightings associated with each of the three types of feedback used against the use of invigilation found 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 less than those using invigilation (Mdn=30.00; $U=341.50$, $Z=-2.334$, $p=0.020$). 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 more than those using invigilation (Mdn=20.00; $U=353.00$, $Z=-2.127$, $p=0.033$). 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,[114] these results appear to support earlier claims [67] that those choosing not to invigilate are placing greater emphasis on formative-style assessment instruments.

[Figure 3 near here]

We are careful to note this does not negate concerns about FAO courses that use no invigilation whatsoever.[76] Indeed, when grouped based on the use of invigilation, no significant differences are found in participants' CCSF scale measures. Aligned with the extra purpose

feedback takes on in FAO courses, we hypothesize that this may be part of an effort at addressing concerns about assessment validity. That is, this feedback is not primarily directed at advancing student learning [25] but at validating students' work and identities through consistent one-on-one contact.[14]

3.4 Further limitations

Related to the use of the feedback framework, two final caveats need to be discussed. First, participant feedback measures were premised on the assumption that assessment weighting (and not simply "assessment") directed student learning. This assumption has received little attention in the current research (see, for example: claims [123] and findings [78]) and therefore needs testing. Second, and perhaps most importantly, the actual benefit to learning of instructor-provided feedback is limited by how that feedback is actually used. As Gill and Greenhow [124] 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, we consider our findings limited to the potential rather than actual effect of feedback and, as a whole, we do not pretend the feedback framework to be a precise but blunt measure of feedback quality.

4. Conclusion

This study focused on the use of assessment and feedback in FAO mathematics courses and asked two broad questions:

1. What assessment instruments are instructors currently using to assess their students?
How are these weighted? Which are invigilated?
2. How are instructors using feedback with these instruments? What kind of feedback?
How is it timed?

Regarding the first question, it is evident that participants use a wide variety of assessment schemes in their FAO courses. Within this variety, there is an overall emphasis on the use of homework and CAA systems as well as, mirroring F2F courses, summative assessment instruments (e.g. exams and tests). However, in seeming contrast with the emphasis in general FAO pedagogy, but consistent with prior FAO mathematics research, we found evidence for a

de-emphasis on student discussion. Furthermore, in seeming contrast with F2F mathematics pedagogy, a large number of course assessment schemes include no invigilation whatsoever and, extending prior findings, we found about one quarter of all mid-term exams and three-quarters of all tests are not invigilated.

Regarding the second question, we found, on balance, that participants emphasize the use of the richer (type 2) assessment feedback, with this feedback most associated with projects and homework, and least associated with final and mid-term exams. However, we found no link between the quality of feedback used and participants' approaches to teaching for conceptual understanding and with a student focus, suggesting this feedback may not be, at least primarily, advancing student learning. Similarly, though results appear to confirm earlier claims that more formative assessment instruments are used by participants who choose not to use any invigilation in their course, this feedback again does not appear primarily directed at advancing student learning. Moreover, further signalling possible areas for future research, we found evidence suggesting shifts in the nature of interactivity in these courses. Specifically, in the context of a reduction in and problems with human interactions, a possible greater reliance on computer-human interactions (e.g. via CAA systems). If this is indeed the case, this raises some concern about how deep and meaningful learning is being achieved in these courses: CAA systems may offer a proxy for needed feedback processes but only one that currently has limited potential for advancing student learning.

The study also provides more generalized findings on assessment practice. To our knowledge, this is the first study to present empirical findings that support claims of blurring with regards to the purposes of assessment instruments; in consideration of their potential impact on learning through the use of feedback, we found that same-named assessment instruments are not equal but use a range of different feedback timings, kinds as well as delivery methods.

Clearly there is a need for more detailed investigations of FAO assessment and feedback practices, and teaching in general, particularly as use of these courses is expected to grow. But there is also an opportunity as this new instructional environment appears to be encouraging experimentation. For example, the use and combination of different assessment instruments suggest this may be an opportunity to research the effects of these changes upon student learning behaviours and the development of their understanding in mathematics. These findings have the potential to inform the undergraduate teaching community as a whole, in both FAO and F2F practice.

References

- [1] Bakia, M., Shear, L., Toyama, Y., & Lassetter, A. (2012). *Understanding the implications of online learning for educational productivity*. Washington, D.C.: U.S. Department of Education Office of Educational Technology Center for Technology in Learning.
- [2] Miller, B. (2010). The course of innovation: Using technology to transform higher education. Washington, D.C.: Education Sector.
- [3] Campbell, M. (2012). Online schooling is exploding in US. *New Scientist*, 215(2881), 6-8.
- [4] Allen, I. E., & Seaman, J. (2011). *Going the distance: Online education in the United States, 2011*. Newburyport, MA: The Sloan Consortium.
- [5] Selden, A. (2005). New developments and trends in tertiary mathematics education: Or, more of the same? *International Journal of Mathematical Education in Science and Technology*, 36(2), 131-147.
- [6] Maltempi, M. V., & Malheiros, A. P. S. (2010). Online distance mathematics education in brazil: Research, practice and policy. *ZDM*, 42(3), 291-303.
- [7] Zinger, V. A. (2006). Challenges and rewards of distance education in Alaska: Experience in teaching web-based courses in mathematics in Alaska. *Proceedings of the 5th IASTED International Conference on Web-Based Education*, Puerto Vallarta, Mexico. 361-364.
- [8] Engelbrecht, J., & Harding, A. (2005b). Teaching undergraduate mathematics on the internet: Part II. *Educational Studies in Mathematics*, 58(2), 253-276.
- [9] Sumler, D. E. (2001). *Distance learning in Maryland colleges and universities, academic year 2000*. Annapolis, Maryland: Maryland State Higher Education Commission.
- [10] Kirkman, E. (2012). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 conference board of mathematical sciences (CBMS) survey*. Providence, RI: American Mathematical Society.
- [11] Glass, J., & Sue, V. (2008). Student preferences, satisfaction, and perceived learning in an online mathematics class. *Journal of Online Learning and Teaching*, 4(3), 325-338.

- [12] Lokken, F. (2011). *2010 distance education survey results - trends in e-learning: tracking the impact of e-learning at community colleges*. Washington, D.C.: Instructional Technology Council.
- [13] Philip, D. (2003). Internet-based tools for mathematics learning: Making math add up on the internet in learning technology innovation in Canada. A supplement to the journal of distance education. *Journal of Distance Education*, 17(3), 38-39.
- [14] Smith, G. G., Ferguson, D., & Caris, M. (2003). The web versus the classroom: Instructor experiences in discussion-based and mathematics-based disciplines. *Journal of Educational Computing Research*, 29(1), 29-59.
- [15] Smith, G. G., & Ferguson, D. (2005). Student attrition in mathematics e-learning. *Australasian Journal of Educational Technology*, 21(3), 323-334.
- [16] Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L. . . . Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, 74(3), 379-439.
- [17] Shachar, M., & Neumann, Y. (2010). Twenty years of research on the academic performance differences between traditional and distance learning: Summative meta-analysis and trend examination. *Journal of Online Learning and Teaching*, 6(2), 318-334.
- [18] Zhao, Y., Lei, J., Yan, B., Lai, C., & Tan, S. (2005). What makes the difference? A practical analysis of research on the effectiveness of distance education. *The Teachers College Record*, 107(8), 1836-1884.
- [19] Davis, L., Harrison, M., Palipana, A., & Ward, J. (2005). Assessment-driven learning of mathematics for engineering students. *International Journal of Electrical Engineering Education*, 42(1), 63-72.
- [20] Houston, K. (2002). Assessing undergraduate mathematics students. In D. Holton (Ed.), *The teaching and learning of mathematics at university level: An ICMI study* (pp. 407-422). The Netherlands: Kluwer Academic Publishers.
- [21] Marriott, P., & Lau, A. (2008). The use of on-line summative assessment in an undergraduate financial accounting course. *Journal of Accounting Education*, 26(2), 73-90.
- [22] Smith, G., & Wood, L. (2000). Assessment of learning in university mathematics. *International Journal of Mathematical Education in Science and Technology*, 31(1),

125-132.

- [23] Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Maidenhead, UK: Open University Press.
- [24] Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: RoutledgeFalmer.
- [25] Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- [26] Taras, M. (2005). Assessment summative and formative: Some theoretical reflections. *British Journal of Educational Studies*, 55(4), 466-478.
- [27] Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4-13.
- [28] Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84.
- [29] Knight, P. T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education*, 27(3), 275-286.
- [30] Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- [31] Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology*, 89(2), 248-262.
- [32] Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- [33] Entwistle, N. (2003). *Concepts and conceptual frameworks underpinning the ETL project*. (No. 3). Edinburgh, UK: Enhancing Teaching-Learning Environments in Undergraduate Courses, Higher and Community Education, School of Education, University of Edinburgh.
- [34] Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- [35] Tobin, K. (1986). Effects of teacher wait time on discourse characteristics in mathematics and language arts classes. *American Educational Research Journal*, 23(2), 191-200.
- [36] Simmons, M., & Cope, P. (1993). Angle and rotation: Effects of different types of feedback on the quality of response. *Educational Studies in Mathematics*, 24(2), 163-

- [37] Zerr, R. (2007). A quantitative and qualitative analysis of the effectiveness of online homework in first-semester calculus. *Journal of Computers in Mathematics and Science Teaching*, 26(1), 55-73.
- [38] Butler, M., Pyzdrowski, L., Goodykoontz, A., & Walker, V. (2008). The effects of feedback on online quizzes. *International Journal for Technology in Mathematics Education*, 15(4), 131-136.
- [39] Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37(1), 57-70.
- [40] Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, 27(6), 501-510.
- [41] van der Kleij, F. M., Theo, J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272. doi: 10.1016/j.compedu.2011.07.020
- [42] Kvale, S. (2007). Contradictions of assessment for learning in institutions of higher learning. In D. Boud, & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (1st ed., pp. 57-71). Oxford, UK: Taylor & Francis.
- [43] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- [44] Entwistle, N. (2009). *Teaching for understanding at university. deep approaches and distinctive ways of thinking*. Basingstoke, UK: Palgrave Macmillan.
- [45] Ramsden, P. (1988). *Improving learning: New perspectives*. London, UK: Kogan Page.
- [46] Anderson, T. D., & Garrison, D. R. (1995). Critical thinking in distance education: Developing critical communities in an audio teleconference context. *Higher Education*, 29(2), 183-199.
- [47] Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32(1), 99-113.
- [48] Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies* (2nd ed.). London, UK:

RoutledgeFalmer.

- [49] Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences*, 2(3), 277-305.
- [50] Roper, W. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology*, 14(1), 43-49.
- [51] Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L., . . . Mulcahy, C. (2011). Feedback during web-based homework: The role of hints. *Artificial Intelligence in Education: Lecture Notes in Computer Science*, 6738(2011), 328-336.
- [52] Baker, S., Gersten, R., & Lee, D. S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, 103(1), 51-73.
- [53] Arend, B. D. (2007). Course assessment practices and student learning strategies in online courses. *Journal of Asynchronous Learning Networks*, 11(4), 3-13.
- [54] Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education*, 21(3), 117-132.
- [55] Liang, X., & Creasy, K. (2004). Classroom assessment in web-based instructional environment: Instructors' experience. *Practical Assessment, Research & Evaluation*, 9(7). Retrieved December 5, 2012: <http://PAREonline.net/getvn.asp?v=9&n=7>
- [56] Robles, M., & Braathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal*, 44(1), 39-49.
- [57] Walker, D. J. (2007). Principles of good online assessment design. *From the Re-Engineering Assessment Practices in Scottish Higher Education (REAP) International Online Conference on Assessment Design for Learner Responsibility, May 29th-31st*, 1-6.
- [58] Anderson, T. (2008). Towards a theory of online learning. In T. Anderson, & F. Elloumi (Eds.), *Theory and practice of online learning* (pp. 45-74). Athabasca, Canada: Athabasca University.
- [59] Jarmon, C. (1999). Testing and assessment at a distance. In M. Boaz, B. Elliott, D. Foshee, D. Hardy, C. Jarmon & D. Olcott Jr (Eds.), *Teaching at a distance: A handbook for instructors* (pp. 55-65). Los Angeles, CA: League for Innovation in the Community College, a Division of Harcourt Brace & Company.

- [60] Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333-2351.
- [61] Shuey, S. (2002). Assessing online learning in higher education. *Journal of Instruction Delivery Systems*, 16(2), 13-18.
- [62] Speer, N. M., Smith III, J. P., & Horvath, A. (2010). Collegiate mathematics teaching: An unexamined practice. *The Journal of Mathematical Behavior*, 29(2), 99-114. doi: DOI: 10.1016/j.jmathb.2010.02.001
- [63] Iannone, P., & Simpson, A. (2011). The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications*, 30(4), 186-196.
- [64] Ross, S. C. (1999). What happened to tests? Assessment practices in undergraduate mathematics. *MAA Notes*, 49, 77-79.
- [65] Solomon, Y. (2007). Not belonging? what makes a functional learner identity in undergraduate mathematics? *Studies in Higher Education*, 32(1), 79-96.
- [66] Wood, L. N., & Smith, G. H. (1999). Flexible assessment. In W. Spunde, P. Cretchley & R. Hubbard (Eds.), *The challenge of diversity* (pp. 229-233). Laguna Quays, AU: University of Southern Queensland Press.
- [67] Trenholm, S. (2007a). An investigation of assessment in fully asynchronous online math courses. *International Journal for Educational Integrity*, 3(2), 41-55.
- [68] Evia, G. B. (2006). Distance learning today. *Proceedings of the International Congress of Mathematicians*, Madrid, Spain. 1752-1757.
- [69] Sakshaug, L. E. (2000). Research on distance education: Implications for learning mathematics. *Focus on Learning Problems in Mathematics*, 22(3/4), 111-124.
- [70] Lingefjård, T., & Holmquist, M. (2002). Mathematics, technology and examination in distance education: Working group 2. *European Research in Mathematics Education II*, Mariánské Lázně, Czech Republic. 193-202.
- [71] Comeaux, P. (2005). *Assessing online learning*. Bolton, MA: Anker Publishing.
- [72] Austin, G. (2007). A review of: 'Assessing online learning'. *Community College Journal of Research and Practice*, 31(1), 79-81.
- [73] Semião, P. (2009). Strategies for a web-based mathematics course. In A. Méndez-Vilas, A. Solano Martín, J. A. Mesa González & J. Mesa González (Eds.), *Research, reflections and innovations in integrating ICT in education* (pp. 436-439). Badajoz, Spain: Formatex.

- [74] Mandernach, B. J., Gonzales, R., & Garrett, A. L. (2006). An examination of online instructor presence via threaded discussion participation. *Journal of Online Learning and Teaching*, 2(4), 248-260.
- [75] Swan, K. (2003). Learning effectiveness online: What the research tells us. In J. Bourne, & J. C. Moore (Eds.), *Elements of quality online education, practice and direction* (pp. 13-45). Needham, MA: Sloan Center for Online Education.
- [76] Trenholm, S. (2007b). A review of cheating in fully asynchronous online courses: A math or fact-based course perspective. *Journal of Educational Technology Systems*, 35(3), 281-300.
- [77] Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15(1), 101-111.
- [78] Wormald, B. W., Schoeman, S., Somasunderam, A., & Penn, M. (2009). Assessment drives learning: An unavoidable truth? *Anatomical Sciences Education*, 2(5), 199-204.
- [79] Allen, G. D. (2003). A survey of online mathematics course basics. *College Mathematics Journal*, 34(4), 270-279.
- [80] Gadanidis, G., Graham, L., McDougall, D., & Roulet, G. (2002). On-line mathematics: Visions and opportunities, issues and challenges, and recommendations. white paper based on the fields institute for research in mathematical sciences mathematics On-line working meeting. Toronto, Ontario: Fields Institute.
- [81] Everson, M. G., & Garfield, J. (2008). An innovative approach to teaching online statistics courses. *Technology Innovations in Statistics Education*, 2(1), October 30, 2012.
- [82] Pomper, M. (2007). Building community in an online upper-division mathematics course. *International Journal of Instructional Technology and Distance Learning*, 4(12), 45-61.
- [83] Tsvigu, C., Breiteig, T., Persens, J., & Ndalichako, J. (2008). Promoting the learning of mathematics: On the use of learning styles in a distance education calculus course. In L. Holtman, C. Julie, O. Mikalsen, D. Mtetwa & M. Ogunniyi (Eds.), *Some developments in research in science and mathematics in -sub-Saharan Africa: Access, relevance, learning, curriculum research* (pp. 133-156). Somerset West, South Africa: African Minds.

- [84] Galante, D. (2002). *Web-based mathematics: An examination of assessment strategies implemented in the online mathematics classroom*. (PhD, Illinois State University).
- [85] Smith, G. G., Heindel, A. J., & Torres-Ayala, A. T. (2008). E-learning commodity or community: Disciplinary differences between online courses. *The Internet and Higher Education*, 11(3-4), 152-159.
- [86] Greenberg, W., & Williams, M. (2008). New pedagogical models for instruction in mathematics. *Mathematical Modeling, Simulation, Visualization and e-Learning: Proceedings of an International Workshop Held at Rockefeller Foundation's Bellagio Conference Center, Milan, Italy*. 361-371.
- [87] Lowe, T., & Hasson, R. (2011). Assessment for learning: Using moodle quizzes in mathematics. *CETL-MSOR Conference 2010*, 39-45.
- [88] Varsavsky, C. (2004). Can online weekly quizzes contribute to learning in mathematics? *Proceedings of the 9th Asian Technology Conference in Mathematics*, Singapore. 161-168.
- [89] Griffin, F., & Gudlaugsdottir, S. (2006). Using online randomised quizzes to boost student performance in mathematics and operations research. *Proceedings of the 7th International Conference on Information Technology Based Higher Education and Training, ITHET 2006*, Sydney, Australia. 21-29.
- [90] Pitcher, N., Goldfinch, J., & Beevers, C. (2002). Aspects of computer-based assessment in mathematics. *Active Learning in Higher Education*, 3(2), 159-176.
- [91] Paterson, J. S. (2002). Linking on-line assessment in mathematics to cognitive skills. *Proceedings of the 6th International CAA Conference*, Loughborough University, Loughborough, UK. 295-304.
- [92] Morrison, G. R., Ross, S. M., Gopalakrishnan, M., & Casey, J. (1995). The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology*, 20(1), 32-50. doi: DOI: 10.1006/ceps.1995.1002
- [93] Hrastinski, S., & Stenbom, S. (2013). Student–student online coaching: Conceptualizing an emerging learning activity. *The Internet and higher education*, 16, 66-69.
- [94] Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20-26.

- [95] Dubinsky, E. (1991). Reflective abstraction in advanced mathematical thinking. *Advanced Mathematical Thinking*, 11(2), 95-126.
- [96] Sangwin, C., Cazes, C., Lee, A., & Wong, K. L. (2010). Micro-level automatic assessment supported by digital technologies. In C. Hoyles, & J. B. Lagrange (Eds.), *Mathematics education and technology-rethinking the terrain* (pp. 227-250). New York, NY: Springer.
- [97] Mascuilli, A. B. (2004). Comparing an online course to its classroom counterpart. *International Journal of Instructional Technology and Distance Learning*, 1(5), 29-31.
- [98] Smith, G. G., Torres-Ayala, A. T., & Heindel, A. J. (2008). Disciplinary differences in E-learning instructional design. *Journal of Distance Education*, 22(3), 63-88.
- [99] Flesch, M., & Ostler, E. (2011, February 28). An analysis of how proctoring exams in online mathematics offerings affects student learning and course integrity. *International Journal for Mathematics Teaching and Learning*. Retrieved December 5, 2012: <http://www.cimt.plymouth.ac.uk/journal/flesch.pdf>
- [100] Yates, R. W., & Beaudrie, B. (2009). The impact of online assessment on grades in community college distance education mathematics courses. *American Journal of Distance Education*, 23(2), 62-70.
- [101] Englander, F., Fask, A., & Wang, Z. (2011). Comment on 'The impact of online assessment on grades in community college distance education mathematics courses' by Ronald W. Yates and Brian Beaudrie. *American Journal of Distance Education*, 25(2), 114-120.
- [102] Trenholm, S. (2013). *Adaptation of Tertiary Mathematics Instruction to the Virtual Medium: Approaches to Assessment Practice*. (Unpublished PhD). Loughborough University, East Midlands, UK.
- [103] Trenholm, S., Alcock, L., & Robinson, C. L. (under review). Assessment and Feedback in Fully Online Tertiary Mathematics: The Instructor Perspective. *Journal for Research in Mathematics Education*.
- [104] Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16(4), 409-424.
- [105] Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the approaches to teaching inventory. *British Journal of Educational Psychology*, 76(2), 405-419.
- [106] Stes, A., De Maeyer, S., & Van Petegem, P. (2010). Approaches to teaching in higher

- education: Validation of a Dutch version of the approaches to teaching inventory. *Learning Environments Research*, 13, 59-73.
- [107] Juan, A. A., Huertas, M. A., Trenholm, S., & Steegmann, C. (2011). *Teaching mathematics online: Emergent technologies and methodologies*. PA, US: IGI Global.
- [108] Radford, A. W. (2011). *Learning at a distance: Undergraduate enrollment in distance education courses and degree programs* (NCES Report No. 2012-154). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- [109] Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195-219.
- [110] Duda, M. D., & Nobile, J. L. (2010). The fallacy of online surveys: No data are better than bad data. *Human Dimensions of Wildlife*, 15(1), 55-64.
- [111] Burn, R. P., Appleby, J. C., & Maher, P. (1998). *Teaching undergraduate mathematics*. London, U.K.: Imperial College Press.
- [112] Engelbrecht, J., & Harding, A. (2005a). Teaching undergraduate mathematics on the internet: Part I. *Educational Studies in Mathematics*, 58(2), 235-252.
- [113] Brookhart, S. M. (2004). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, 2004(100), 5-14. doi: 10.1002/tl.165
- [114] Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.
- [115] Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21-27. doi: 10.1016/j.stueduc.2012.04.001
- [116] Kirkman, E., Lutzer, K. J., Maxwell, J. W., & Rodi, S. B. (2007). *Statistical abstract of undergraduate programs in the mathematical sciences in the united states: Fall 2005 conference board of mathematical sciences (CBMS) survey*. Providence, RI: American Mathematical Society.
- [117] Finnegan, C., Morris, L. V., & Lee, K. (2009). Differences by course discipline on student behavior, persistence, and achievement in online courses of undergraduate general education. *Journal of College Student Retention: Research, Theory and Practice*, 10(1), 39-54.
- [118] Illowsky, B. S. (2007). *Effects of discussion postings in online elementary statistics community college classes*. (PhD, Capella University).

- [119] Lowrie, T., & Jorgensen, R. (2012). Teaching mathematics remotely: Changed practices in distance education. *Mathematics Education Research Journal*, 24(3), 371-383.
- [120] Anderson, T., & Elloumi, F. (2008). *The theory and practice of online learning*. Alberta, Canada: AU Press.
- [121] Moore, M. G. (2007). The theory of transactional distance. In M. G. Moore, & W. G. Anderson (Eds.), *Handbook of distance education* (2nd ed., pp. 89–104). Mahwah, NJ: Lawrence Erlbaum.
- [122] Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six online statistics courses: Examination and review. *The American Statistician*, 59(3), 240-251.
- [123] Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London, UK: Kogan Page Ltd.
- [124] Gill, M., & Greenhow, M. (2008). How effective is feedback in Computer-Aided assessments? *Learning, Media and Technology*, 33(3), 207-220.

Tables with captions:

Table 1: A proposed taxonomy of kinds of feedback in relation to the effect on learning

Description of Feedback	Quality of Feedback	Related Basis in the General Literature	
		Claims	Empirical Studies
Correct/Incorrect	Poor	[40, 42, 43]	[41]
Full solution	Intermediate	[45 as cited in 46]	[30, 47, 51]
Hints or comments	Rich	[24, 30, 44, 48, 49]	[25, 47]

Table 2: Assessment and feedback specifics questions for each identified assessment instrument

1. Type (select one only):						
<i>Homework</i>	<i>Quizzes</i>	<i>Tests</i>	<i>Mid-term exam</i>	<i>Final exam</i>	<i>Individual project</i>	<i>Group project</i>
<i>Portfolio</i>	<i>Group work (e.g. problem solving)</i>	<i>Journal</i>	<i>Discussion participation-QUANTITY</i>	<i>Discussion participation-QUALITY</i>	<i>Discussion participation-BOTH</i>	<i>Other</i>
2. Weighting (“50” if worth 50% in total):						
3. Number (e.g. “4” if four tests):						
4. Author (select all that apply):						
<i>Self</i>	<i>Group (e.g. department)</i>		<i>Commercial (e.g. textbook)</i>		<i>Other</i>	
5. Question types (select all that apply):						
<i>Multiple-choice</i>		<i>Short answer</i>		<i>Long answer</i>		<i>Other</i>
6. Supervision? (select one only)						
<i>Invigilated</i>		<i>Not invigilated</i>			<i>Other</i>	
7. Constrained by Department or Institution (select all that apply)						
<i>Weighting</i>		<i>Number</i>		<i>Invigilation</i>		<i>Other</i>
8. Typical timing to feedback (e.g. immediate, 1 wk, 24 hrs...):						
9. Kind of feedback visible to the student? (please select ALL that apply):						
<i>Grade/Mark</i>						
<i>Correct answer provided</i>						
<i>Computer-generated full solution</i>						
<i>Lecturer-generated full solution</i>						
<i>Computer-generated hints or comments</i>						
<i>Lecturer-generated hints or comments</i>						
<i>Hints or comments challenge student understanding</i>						
<i>Peer</i>						
<i>Other (specify):</i>						

Table 3: Online Survey Participants and Institutional Context

Country/Group	“2 yr” (e.g. US community college)	“4 yr” (e.g. US state university)	“Online” (e.g. Open University)	Other	Unclear from response	Total	Percent (1 d.p.)
US (SLN)	25	7	0	3 ^b	2	37	52.9%
US (RODP/RUME)	6	10	0	0	-	16	22.9%
IGI Global ^a	0	7	0	1	-	8	11.4%
Other	0	2	7	0	-	9	12.9%
TOTAL	31	26	7	4	2	70	100%

^a Publisher for the research compilation.^b A combination of 2 and 4 yr institution, as specified by participants.

Table 4: A framework, based on claims and findings discussed in the literature review, for distinguishing assessment instrumentation according to kind of feedback provided.

Operationalization of the Framework for Feedback						
Quality of Feedback	Poor ←			→ Rich		
Kind of Feedback as Described by the Survey Instrument	Grade/Mark	Correct answer provided	Computer-generated full solution/ Instructor-generated full solution	Computer-generated hints or comments	Instructor-generated hints or comments	Hints or comments challenge student understanding
Feedback Type/Score	0	1		2		

Table 5: Assessment instruments used with descriptive statistics

Instrument	Total Number Using (of n=66)	Weighting (%)	
		Mean	Standard Deviation
Homework	55	28	26.4
Final Exam	48	33	20.3
Tests	43	42	22.9
Quizzes	35	20	9.4
Discussion	26	10	5.3
Mid-Term Exam	16	23	12.4
Individual Project	13	14	4.7
Other Assessment ^a	8	13	9.9
Group Project	3	13	5.8
Group Work	2	5	7.1
Journaling	1	10	-
Portfolio	1	5	-

^a 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.

Table 6: Use of invigilation with instruments

Instrument	Not Using Invigilation			Using Invigilation		
	Percent of total number using specified instrument	Actual Number	Avg. Weighting (%)	Percent of total number using specified instrument	Actual Number	Avg. Weighting (%)
Homework	100	50	27.7	0	0	-
Quizzes	100	32	20.2	0	0	-
Discussion	95	21	9.6	5	1	20
Individual Project	92	11	13.0	8	1	10
Other Assessment	75	6	14.8	25	2	5.5
Tests	71	27	36.1	29	11	56.8
Final Exam	27	12	31.4	73	33	34.0
Mid-Term Exam	27	4	17.2	73	11	25.6

Table 7: Quality of feedback as measured by the study framework

Instrument	Total Number Using (of n=66)	Average Feedback Score (2 d.p.)	Number of Participants Using each Type of Feedback		
			Type 0	Type 1	Type 2
Homework	55	1.73	1	13	41
Final Exam	48	0.52	30	11	7
Tests	43	1.23	8	17	18
Quizzes	35	1.26	1	24	10
Discussion	26	1.00	12	2	12
Mid-Term Exam	16	0.94	5	7	4
Individual Project	13	1.85	0	2	11
Other Assessment	7	1.50	1	2	5

Table 8: Timing of feedback

Instrument	Total Number Using (of n=66)	Frequency (Mean)	Timing of Feedback		
			Days		Missing Data ^a
			Mean	S.D.	
Homework	55	14.9	1.8	3.8	3
Final Exam	48	1.1	4.8	6.5	9
Tests	43	6.3	2.2	2.6	2
Quizzes	35	9.9	0.8	2	2
Discussion	26	8.9	4.6	4.8	3
Mid-Term Exam	16	1.4	4.5	2.9	1
Individual Project	13	4	5.8	4.5	1
Other Assessment	7	5.7	4.3	3.4	1

^a a number of participants failed to identify the timing of their feedback for the specified instrument.

Table 9: Use of Computer-generated Feedback

Instrument	Participants Using Computer-generated Feedback		Weighting (%)		
	Of Total Number	%	Mean		CAA Standard Deviation ^a
			CAA	Non-CAA	
Homework	34/55	62	18	44	16
Final Exam	5/48	10	27	33	5.2
Tests	14/43	33	44	43	26.9
Quizzes	27/35	77	20	21	9.7
Discussion	3/26	12	10	10	
Mid-Term Exam	1/16	6	7	25	
Individual Project	0/13	0	-	14	
Other Assessment	2/8	25	19	10	

^a Due to low numbers, standard deviations are not considered for the last four instruments.

Table 10: Comparison of US participants' use of CAA with 2005 US statistics [116]

US Institution Type	Percent Using CAA		
	2005 F2F Sections Using Online Homework and/or Testing/Grading [116]	2010 FAO Courses Using CAA Homework ^a [102]	Difference
2yr ^b	5.3%	65% (22/34)	+59.7%
4yr ^c	6.3%	61% (11/18)	+54.7%

^a 2010 numbers reflect US-only data. ^b Average across 28 categories of courses compared with 17 in present study. ^c Elementary statistics only (2.2% for Calculus I and II) compared with 13 of 18 in the present study which were either elementary statistics or pre/calculus level course.

Figure captions:

Figure 1: 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)

Figure 2: Individual participants' assessment practice by type of feedback used

Figure 3: Use of invigilation related to individual participants' assessment practices (type of feedback used)

Figure 1

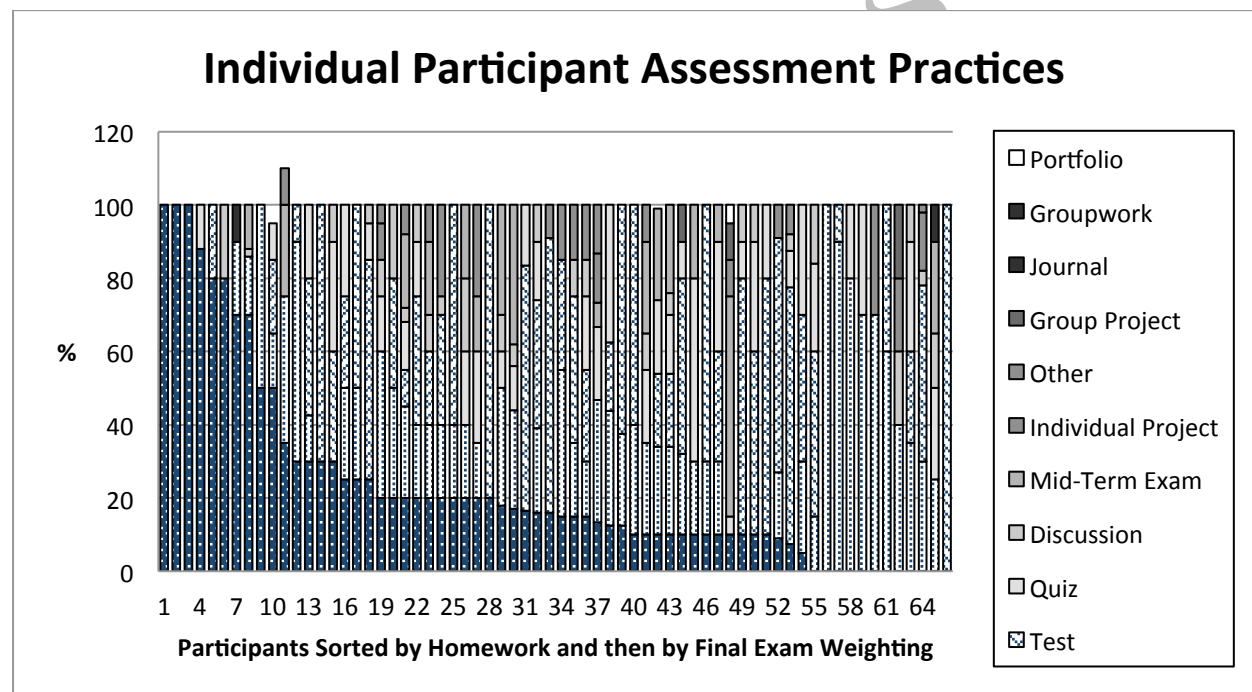


Figure 2

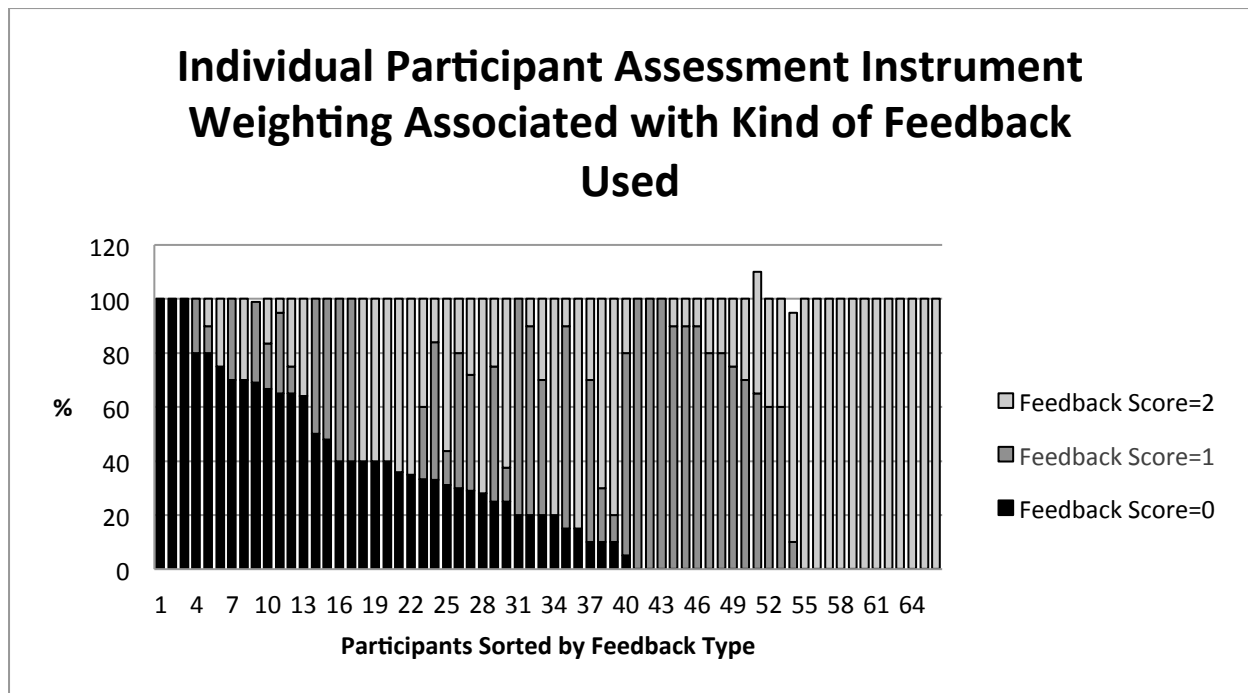


Figure 3

