2017-07-01

Utility of Feedback Given by Students During Courses

Michael Alton Atkisson
Brigham Young University

Utility of Feedback Given by Students During Courses

Michael Alton Atkisson

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

Richard E. West, Chair
Charles R. Graham
Daniel T. Hickey
Ross Larsen
Stephen C. Yanchar

Department of Instructional Psychology and Technology

Brigham Young University

ABSTRACT

The Utility of Feedback Given by Students During Courses

Michael Alton Atkisson
Department of Instructional Psychology and Technology, BYU
Doctor of Philosophy

This two-article dissertation summarizes the end-of-course survey and formative feedback literatures, as well as proposes actionability as a useful construct in the analysis of feedback from students captured in real-time during their courses. The present inquiry grew out of my work as the founder of DropThought Education, a Division of DropThought. DropThought Education was a student feedback system that helped instructional designers, instructors, and educational systems to use feedback from students to improve learning and student experience. To find out whether the DropThought style of feedback was more effective than other forms of capturing and analyzing student feedback, I needed to (1) examine the formative feedback literature and (2) test DropThought style feedback against traditional feedback forms. The method and theory proposed demonstrates that feedback from students can be specific and actionable when captured in the moment at students' activity level, in their own words. Application of the real-time feedback approach are relevant to practitioners and researchers alike, whether an instructor looking to improve her class activities, or a learning scientist carrying out interventionist, design-based research.

Key words: formative feedback, end-of-course feedback, real-time feedback, DropThought, hierarchical generalized linear model, text classification

ACKNOWLEDGMENTS

I dedicate this work to my parents, my wife, and my children. To my mother, who saw me start my Ph.D. work, but did not see me finish, I am particularly indebted. She raised me, loved me, and unfortunately for her, edited many of my papers. To my father, whose patience and support has helped keep me on the path. To my dear wife, who was with dealt half a husband for the past 9 years, and yet was supportive and loving to the end. To my precious children for the countless dad hours sacrificed in the name of finishing what I started. It takes a village, and this dissertation is ours more than mine.

I would like to give special thanks to Dr. Richard E. West, my chair. I could not have done this work without his superhuman level of patience and persistence. Dr. West is the second author on Article 1 because of his thorough feedback on the argumentation, content, and syntax of the article. He is third author on Article 2 for the same. Dr. Ross Larsen helped me reason through my analysis and methods and gave detailed reviews on my descriptions of the method and results in the Method, Results, and Discussion sections of Article 2, without whom I would have been lost. Accordingly Dr. Larsen in second author on Article 2. Dr. Charles Graham and Dr. Stephen Yanchar kept me honest in my claims and reasoning in the prospectus and dissertation review. Dr. Daniel Hickey helped inspire me to pursue scholarship and helped ground my work in defensible theory in the prospectus and dissertation reviews.

Other mentors have helped me greatly along the way. Dr. David Wiley helped me discover learning analytics and MOOCs as my first committee chair. Thank you, David, for helping me dream big. Dr. Joseph Olsen in Brigham Young University's Sociology department was critical to helping me learn to reason with statistics. Dr. Thomas Schwen captured my imagination with the social and cultural aspects of learning so much that I followed him to

Indiana. He was a thoughtful thought provoking chair. Dr. Stephanie Allen was critical to my joining the education and instructional design world. She guided me early on to tackle learning problems with practical significance in the world.

As I have worked full-time throughout my doctoral program, several supervisors and employers burdened themselves with taking on a part-time Ph.D. student. Dr. Michael Nobel at Allen Communication Learning Services enabled me to consolidate and reduce hours so that I could begin my course work. Dr. Joseph South hired me at K12, Inc. to help re-envision graduate business education while I was taking classes and beginning my research at BYU. Dr. Sangil Yoon, a fellow Indiana University Bloomington Instructional Systems Technology alumnus, hired me at the Graduate School of Business at Stanford while I performed MOOC research on the side. Karan Chaudhry hired me at DropThought amid my research even though time, funds, and resources were tight at the start up. Beth Davis and Russ Little hired me just after the PAR Framework was acquired by Hobsons while I completed my prospectus. Howard Bell, our new general manager at Hobsons has given me considerable leeway to finish the dissertation. I could not have completed this work without the faith and confidence of these individuals and the many team members who supported me at these companies. Thank you to all.

Also, a special thanks to David Atkisson, who wrote a Python script to split student comments on to different rows in the comma delimited file that contained my data. Wherever I placed a special character, the script would separate the remaining comment text on a separate row the while carrying down and incrementing meta data in adjacent columns. With over 2,000 comments, this saved many hours.

Lastly, a special thanks to Dr. Deb Adair at Quality Matters for believing in real-time feedback and putting it on the map. Deb was critical in allowing this research to exist as she gave DropThought Education its first big break and approved the use of Quality Matters data for this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

Article 2 Page

Description of Research Agenda and Structure of Dissertation

Higher education in the United States (U.S.) is faced with many challenges today. The most recent data from the Institute of Education Sciences ("The NCES Fast Facts Tool," 2015) shows that the 6-year graduation rate in 2013 for 4-year institutions was 59 percent for first-time, full-time students. Three-year U.S. junior college graduation rates in 2013 were even worse at 29.4 percent for first-time, full-time students ("Digest of Education Statistics," 2014). Over the past 30 years, tuition at 2-year colleges has increased on average 3.1% per year over inflation, and published tuition prices have increased by 146% at private 4-year institutions, 150% at 2-year public institutions, and 225% at 4-year state institutions for in-state students (College Board, 2015). Student debt was recently reported to be as high as $1.2 trillion dollars (Lorin, 2014). Stakeholders from school administrators to instructors face a crisis in being able to help more students succeed in higher education, and the many who are not helped can face devastating, lifelong financial consequences.

The status quo for U.S. higher education is insufficient for nearly 40% of 4-year and about 70% of 2-year full-time students, and part-time students are even more disadvantaged. Despite U.S. higher education having the most educated of any workforce, schools are ill-equipped to handle these challenges. This is evident by 1 out of 3 students not returning for a second year of college each year (U.S. News, 2015). Many aspects of current educational practices may contribute to the disengagement students feel. For example, students face several disadvantages in large freshmen class lecture halls such as being less active in the learning process and having less frequent quality interactions with instructors (Cuseo, 2007). Students in large classes also experience less in-depth thinking in class, less breadth and depth of subject matter, less academic achievement, and less overall course satisfaction (Cuseo, 2007) than they would if in smaller, more personal classes.

Facing this great performance gap in higher education, many schools have turned to predictive analytics to find ways of increasing student success. Various companies provide solution platforms including Education Advisory Board, Inside Track, Ellucian, and Civitas Learning. Academia has also focused on ways to harness big data in education with the Society of Learning Analytics Research (SoLAR) in its fifth year ("SoLAR," 2015) and the International Society of Educational Data Mining (EDM) in its eighth year ("International Educational Data Mining Society," 2015). SoLAR focuses on human-led data exploration and interpretive methods, while EDM focuses on automation in data exploration (Baker & Siemens, 2014). Though SoLAR and EDM differ in their approach, they both rely on large, structured datasets to derive insights and predictive power. Data from student demographic databases or learning management system (LMS) logs are common sources (Baker & Siemens, 2014).

Though large, structured datasets may be effective at affording predictions, predictions do not bring educators closer to understanding phenomena. As Philip Frank noted,

> Scientific findings (validated predictions or observations) outstrip the common sense understanding of them, taking us back to that condition earlier in history where we could control and predict without knowing why, what, or how such regularities in events were really brought about. Man predicted his course of travel under the stars, controlled the crops through practical know-how, and cured himself of certain diseases centuries before there was anything like a scientific account of these beneficial outcomes. (Rychlak, 1988, p. 169)

Extending on this point, Atkisson and Wiley (2011) explained,

> mathematical models, control, and prediction are not sufficient to answer questions about why something happens or what it means. Furthermore, given the irreducibly interpretive nature of inquiry, not attempting to answer questions of meaning and purpose may easily lead to the wrong conclusion, even if one is able to replicate observed behavior . . . While these fishing expeditions may uncover seemingly interesting relationships between constructs, without an interpretive framework grounded in specific theoretical commitments, the data tail may come to wag the theory dog. (p. 119)

The challenge with large, structured datasets in education is that even though they allow for modeling of individual cases, they only result in decontextualized, impersonal data abstractions mapped to individuals. Such results only provide probabilities, not answers to why and how questions for the agents in question (Atkisson & Wiley, 2011) or for those trying to help them.

Education is not in need of more big data, but rather, more—better data. Better data may be found in text data that students leave behind through their student experience. This data is not hard to find, it is just difficult to capture, organize, and semantically associate through automated means. Unstructured text data from students has been available since students began writing essays. However, only with the advancement of digital and mobile technologies has unstructured text data from students been available in a form that lends itself to large-scale analysis. Computational linguistics and Natural Language Processing (NLP) began as rule-based methods for modeling semantics in language and translation (Hutchins, 1999; Martin & Jurafsky, 2000). More recently they have developed into algorithm-based clustering or statistical modeling of language, such as sentiment analysis (Pang & Lee, 2008) and latent semantic analysis (Dumais, 2004). Essentially these NLP techniques enable researchers to identify topics, semantic patterns, and sentiments among the text data exhaust that students leave behind in today's technology mediated learning spaces.

Now it is possible to use automated, predictive analytics to identify, not only which students may be at risk, but also why they may be at risk from their own words captured from various learning episodes they experienced. To the extent that such language modeling is automated, classification of student text can be aligned in real-time to semantically grouped categories, meaningful to efforts in student success or learning. Such mass-identification of individualized summaries of experiences has the potential to usher in a new era of personalized student support. Given the need to make higher education more effective for at-risk students, tools are needed that enable the most profound effects for the most people.

The promise that text data has for transforming student experience into affordances for student progress is the common assumption behind the 2 articles in this dissertation. In the first article, "A Review of Formative Student Feedback Practice and Supporting Tools," I establish a framework for the instructor practice of facilitating students' formative feedback on a course and evaluate the tools that can support this evaluative practice. I anticipate submitting this article to a journal that accepts reviews of literature in technology or data analytics, such as *Educational Technology Research and Development* or *Journal of Educational Technology and Society.*

In the second article, "Towards Actionable Course Design Data through Real-time Feedback," I explore the nature of student feedback captured real-time throughout a course and examine whether feedback captured from students throughout a course is more actionable for course design decisions than feedback captured at the end of the course. I anticipate submitting this article to *Computers and Education, Internet and Higher Education, Educational Technology and Society* or a related top-journal in educational technology.

ARTICLE #1: A REVIEW OF FORMATIVE STUDENT FEEDBACK PRACTICE AND SUPPORTING TOOLS

Running head: Formative Student Feedback Practice and Tools

A Review of Formative Student Feedback Practice and Supporting Tools

Michael A. Atkisson and Richard E. West

Brigham Young University

Abstract

Despite widespread use of end-of-course surveys and the arguments about the validity and reliability of their resulting data, the post-facto feedback method shows no measurable effect on the improvement of courses and instruction over time. Formative feedback, on the other hand, (from students to instructors) has shown promising results in both helping instructors reflect on their courses and teaching, as well as showing improvement of instructor ratings over time. Mobile and social technologies have enabled formative approaches to student feedback to grow significantly in recent years. In this literature review article, we examined the literature for the steps that instructors take to collect, analyze, and make use of formative feedback they capture from students. We also identified the tools used for formative feedback in the academic literature and compared their affordances with the literature's recommended feature sets. The analysis resulted in a Framework called the Lifecycle of Student Feedback. Using this framework, we found considerable gaps in formative feedback tool support for critical steps in the formative feedback process.

Introduction

Student feedback through end-of-course surveys is the most common source of data available to instructors about their performance and that of their courses (Abrahams & Friedman, 1996; Berk, 2005). Student evaluation of teaching is one of the most widely researched topics in education (Cashin, 1999; McKeachie & Kaplan, 1996; Theall & Franklin, 1990), with thousands of research articles in the literature. Several reviews have been done (Aleamoni, 1999; Annan, Tratnack, Rubenstein, Metzler-Sawin, & Hulton, 2013; Braskamp & Ory, 1994; Clayson, 2009; Marsh, 1984, 2007; Marsh & Roche, 1997; Shao, Anderson, & Newsome, 2007), largely demonstrating that student ratings are reliable and valid measures of instruction. For example, Aleamoni's (1999) review of 16 myths regarding student ratings studies from 1924 to 1998 showed, among several other results, that students reliably rated instructors year to year (correlations between .87 and .89) and from course section to section (correlations between 0.70 and 0.87). Aleamoni also found that across 16 studies, student ratings correlated at a moderate to a high level with instructor peer ratings, expert judge ratings, alumni ratings, and student learning measures.

Nevertheless, many faculty remain skeptical of the value that student ratings provide (Beran, Violato, & Kline, 2005; Franklin & Theall, 1989; Nasser & Fresko, 2002; Schmelkin, Spencer, & Gellman, 1997), and many researchers have found significant biases in the post-facto ratings that students provide, including bias ratings for minority status of face-to-face instructors (Carle, 2009), perceived attractiveness (Ambady & Rosenthal, 1993), grade leniency (Greenwald & Gillmore, 1997), age (Wilson, Beyer, & Monteiro, 2014), and gender (Basow, 1995). The few longitudinal studies on the impact of student ratings on instruction show that ratings alone have no effect on the improvement of instruction over time (Lang & Kersting, 2007; Marsh, 2007; Stevens & Aleamoni, 1984).

However, recent research on formative feedback from students to instructors (formative student feedback) throughout the school term has shown promise in affecting instructor reflection and improvement of instruction (Winchester & Winchester, 2011a, 2014). The effectiveness of formative student feedback suggests that the cadence, or the intervals when feedback is given, has a significant impact on instructors' reflection practices and the impact they have on the course. Student evaluations of instruction are also sometimes administered during the course, and referred to as "formative" evaluations, as they can help effect change before the course is completed, but they usually resemble the end-of-course survey and are administered only once (Gravestock & Gregor-Greenleaf, 2008). For the purposes of our review, mid-term evaluations are not considered formative student feedback.

With the recent development of digital and mobile technologies that facilitate quick and easy feedback interactions between students and instructors (Stieger & Burger, 2010), the literature on formative student feedback is growing, and in particular the variety of tools used to facilitate formative student feedback. Given the promise of formative feedback to influence instructional effectiveness and the near ubiquity of digital and mobile devices that support formative feedback interactions, a review of the tool affordances that best facilitate the practice of reflective instruction is merited and should focus on (1) a description of formative feedback practice, and (2) how formative student feedback tools align with the suggested practice.

Hence, the questions for this review are the following:

1. What are the formative feedback practices that instructors use to capture and act on student feedback throughout a school term?
2. How do the reviewed formative feedback tools support formative feedback practice?
3. What affordances should be included in effective formative feedback support tools?

Method

We searched for studies in which instructors, researchers, or institutions captured student course feedback to instructors at a regular, continuous, or ad hoc cadence throughout the school term. See Table 1 for the full list of search terms organized by topic. Though every combination of all terms in Table 1 was not searched, when terms yielded large numbers of results, additional terms were used as qualifiers to narrow the pool of articles. For example, there were no articles on feedback when searching for tools like Facebook, Yik Yak, and Yammer. So, additional qualifiers in those cases were not needed. Whereas with Twitter, various results were found, so it was important to specify other terms such as student evaluation of teaching (SET), student ratings, student feedback, instructor feedback, or student satisfaction. We used quotes around phrases to specify nuances in the search to be sure we were exhaustive in certain search results. For example, end-of-course survey vs. end-of-course evaluations. We also used certain words that were part of phrases in other searches. For example, we searched for formative evaluation as a phrase, but we also searched for formative with other terms like feedback or student feedback. As we came across tool names in articles we searched for them exhaustively, but we also searched for tools that had related features. For example, Facebook has groups like Twitter, but it turned out that there were no articles about student feedback facilitated by Facebook. We began by searching for key terms in Google Scholar, and then proceeded to the following database collections in order to find additional sources: Association for Computer Machinery, EBSCO Host, JSTOR, Proquest, Sage, Science Direct, Springer, T and D Online, and Wiley Online Library.

Lastly, because *student feedback* is a widely-used term that references educational practices that often vary between students giving and receiving feedback, it became necessary to use and qualify several search terms. For example, exclusion of *formative assessment* was key to avoiding articles that focused on the feedback direction of instructor to students. Other exclusions included peer-student feedback, and peer-instructor. Next, we used Google Scholar to search the bibliographies of key articles in order to identify additional studies.

The effort resulted in 100 articles, 10 of which exhibited tools for the facilitation of formative feedback from students to instructors. Peer reviewed articles were the primary focus of the search, though scholarly books and handbooks, conference presentations, and dissertations were included when articles were lacking in particular areas of the search or when the non-peer reviewed sources were highly cited or cited by key articles. Though recent technology tools on the market that can facilitate student-to-instructor feedback were queried in the literature such as Facebook, Yik Yak and others, Twitter was the only on-the-market tool that resulted in articles meeting the criteria. The remainder of the articles about tools discussed general technologies such as online surveys and email or in-house built technologies not available to the public.

We then identified the steps that instructors and researchers took themselves or recommended in formative student feedback practice literature. We also identified the properties of student feedback that were associated with effective formative feedback practice. We labeled each finding by topic and subtopic, which we then analyzed descriptively looking at overall numbers. As similar categories arose we consolidated them. We grouped these steps into a framework of 7 themes. We divided the 7 themes into 4 categories by the instructor's behavioral role: Inquirer, Instructor, Evaluator, and Interventionist. We then reviewed each article to identify recommendations or descriptions about student feedback results and process.

Table 1

Literature Search Terms for Formative Feedback and Tools by Category Type

Tool	Course Evaluation	Cadence	Other Qualifiers	Exclusions
Microblogging	Course evaluations	Formative	Use, Usage	Assessment
Twitter	End-of-course evaluations	Formative evaluation	Instructor-prompted	Problem-based
Yammer	End-of-course survey	Continuous, Real-time	Help-seeking	Formative assessment
Facebook Yik Yak	Student evaluation of teaching (SET)	Ad hoc	Interactive teaching	Feedback to student
	Student ratings	Repeated measures	Class Communication	Peer
Snapchat	Student feedback	Intensive repeated measures	On learning	
	Instructor feedback		Quality assurance	
	Student satisfaction		Reflective practice, teaching	
	Course satisfaction		Epistemology of	
			Review of	

End-of-Course Student Feedback

Capturing student feedback is a common practice in higher education today and is almost entirely carried out through end-of-course surveys (Stieger & Burger, 2010). As Berk (2005) noted from a 1991 U.S. Department of Education newsletter, 98% of 40,000 department chairs in higher education used student evaluations to assess instruction. End-of-course survey results are mostly used for summative evaluation of the effectiveness of instruction (Stieger & Burger, 2010) in various higher education processes including tenure review (Emery, Kramer, & Tian, 2003) and accreditation reporting (Estelami, 2015).

Despite their ubiquity and regularity, however, end-of-course surveys appear not to have had an impact on the increase of instructor effectiveness over time (Lang & Kersting, 2007; Marsh, 2007; Stevens & Aleamoni, 1984). Lang and Kersting (2007), for example, studied 12 instructors over 2 years at an institution that offered no augmentation or help to instructors for the improvement of instruction and previous to the study had not implemented end-of-course surveys. Over 3,000 questionnaires across four consecutive semesters were examined, and Lang and Kersting (2007) found an initial jump between semesters one and two from a mean rating of 2.24 and standard deviation (SD) of 0.13 to a mean rating of 2.34 (SD 0.15). Subsequently, however, Lang and Kersting (2007) noted a steady decline for the remaining 2 semesters in mean ratings, 2.29 (SD 0.14) and 2.25 (SD 0.13) respectively. Stevens and Aleamoni (1984) also collected student feedback at a school that had no history of student rating surveys, but continued the study over 10 years and found that time had no effect on student ratings. Similarly, Marsh (2007) found at a school without a zero baseline for student evaluations of instruction, that among 6,024 undergraduate- and graduate-level university courses taught by 195 instructors over a 13-year period, no improvement or decline in instruction was observed over time. The literature suggests that after initial increases in student ratings at the beginning of observation, there is little evidence for instructional improvement over time due to student evaluations alone, suggesting additional support, training, professional development, and perhaps more useful feedback information is needed to improve teaching.

Formative Student Feedback Practice

Seeing the need to effectively capture more granular feedback on courses and their instruction, several researchers have turned to formative student feedback (Aultman, 2006; Desai, 2014; Goldfarb & Morrison, 2014; Hendry, Cumming, Lyon, & Gordon, 2001; McKone, 1999; Ravelli, 2000; Wagner et al., 2015; Winchester & Winchester, 2010, 2011a, 2012). The advantages of formative feedback over summative feedback for instructors includes increased volume and detail in feedback (Desai, 2014), greater reliability in ratings (Goldfarb & Morrison, 2014) and the opportunity to make changes to teaching (Ravelli, 2000; Winchester & Winchester, 2011a). Students giving the formative feedback have also been observed anecdotally to experience a variety of benefits, including timeliness of interventions (Aultman, 2006), increased student satisfaction from seeing instructor responsiveness (Aultman, 2006; Hendry et al., 2001; Ravelli, 2000), increased student engagement (Aultman, 2006), and increased ownership of learning (Aultman, 2006; Ravelli, 2000). Furthermore, exploratory research (Winchester & Winchester, 2014) on the impact of formative feedback to instructors showed that the deeper level of reflection exhibited by instructors in response to formative student feedback, the more increase those instructors received in effectiveness of instruction ratings. These findings support the idea that instructor reflection is more effective throughout a course as opposed to at its end, because instructors would have an easier time recalling situations mentioned in the feedback (Winchester & Winchester, 2014). As Ramsden said, "Evaluation at the end of a course, cannot replace evaluation during it" (2003, p. 242).

Given the growing practice of formative feedback and the recognition of its effectiveness, a closer look at how instructors and researchers carry it out is merited. This leads to the first research question: What are the formative feedback practices that instructors use to capture and act on student feedback throughout a school term?

The Lifecycle of Student Feedback steps are a curation of recommend formative feedback practice. In answer to research question one, the following summarizes the themes. Subsequently, the remainder of the review describes the themes in detail.

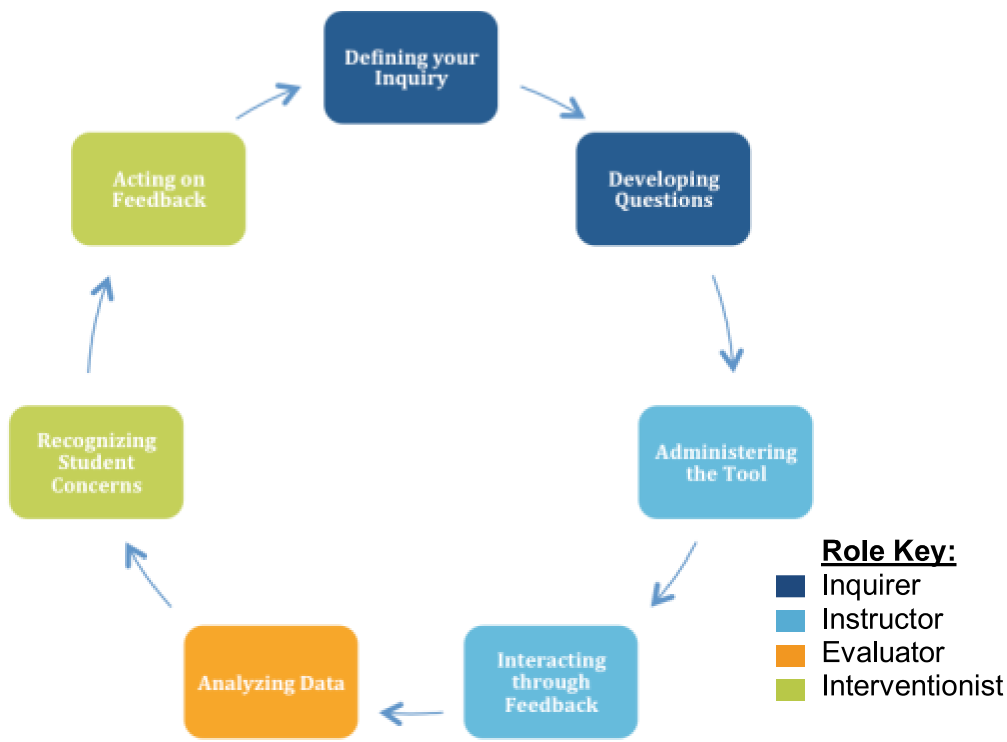


Figure 1. *Lifecycle of Student Feedback by instructor steps and role*

Table 2

Feedback Practices by Instructor Action/Property and Reference

Practice	Action/Property	Reference
Defining your inquiry	Personal theory of teaching	(Biggs & Tang, 2011)
	Clarification of inquiry goals	(Weston, McAlpine, & Bordonaro, 1995)
	Formative	(Lewis, 2001)
	Reflect instructor's personal goals	(Fuhrmann & Grasha, 1983) as cited by (Lewis, 2001)
Developing questions	Iteration of varied questions	(Aultman, 2006; Mosteller, 1989; Ravelli, 2000)
	Broadly-based questions	(Fuhrmann & Grasha, 1983) as cited by (Lewis, 2001)
	Qualitative	(Fuhrmann & Grasha, 1983; Harvey, 2003; Jara & Mellar, 2010; Lewis, 2001; Ravelli, 2000; Watson, 2003)
	Open-ended and specific	(Hendry et al., 2001; Tulgan, 1999; Wagner et al., 2015)
Administering the tool	Quick and easy	(Beaty, 1997; Desai, 2014; Stieger & Burger, 2010)
	Minimal questions	(Desai, 2014; Stieger & Burger, 2010)
	Frequent or continuous	(Fuhrmann & Grasha, 1983; Goldfarb & Morrison, 2014; Harvey, 2003; Jara & Mellar, 2010; McKone, 1999; Tulgan, 1999; Wagner et al., 2015)
	Avoid real-time display	(Fabris, 2015)
Interacting through feedback	Anonymous	(Dennen & Bonk, 2007; Goldfarb & Morrison, 2014; Hendry et al., 2001; Svinicki, 2001; Wagner et al., 2015; Watson, 2003)
	Timely and ad hoc	(Hendry et al., 2001; Lewis, 2001; Tulgan, 1999; Wagner et al., 2015)
	Bidirectional	(Mosteller, 1989; Tulgan, 1999; Wagner et al., 2015)

Table 3

Feedback Practices by Evaluator/Interventionist Action/Property and Reference

Practice	Action/Property	Reference
Analyzing and making sense of data	Summarize data	(Stieger & Burger, 2010)
	Aggregate data	(Stieger & Burger, 2010)
	Facilitate quick, repeated use	(Foth, Fitz-Walter, Ti, Russell-Bennett, & Kuhn, 2012; Knol, 2013; Stieger & Burger, 2010; Winchester & Winchester, 2011a)
Recognizing student concerns	Share analyzed feedback with student in next course session	(Foth et al., 2012; Mosteller, 1989)
	Share with students periodic, anonymous feedback summaries	(Foth et al., 2012; Wagner et al., 2015)
Taking action with students and courses	Follow up structure	(Goldfarb & Morrison, 2014)
	Share resulting actions with students	(Crews & Curtis, 2011; Hendry et al., 2001; Svinicki, 2001; Watson, 2003)

Inquirer

Instructors act in the Inquirer role by deciding what to investigate with formative student feedback and writing the questions to ask the students. First, instructors must decide what to ask on singular occasions and on an ongoing basis. As Weston et al. (1995) suggested, it's a design process. The design process, however, must be *theory-based* in order for actions to be obvious as a result of the feedback. Acting without a theory base results in minor course alterations rather than meaningful change through reflective instructional practice (Biggs, 2001; Bowden & Marton, 2003) as cited by (Boerboom, Stalmeijer, Dolmans, & Jaarsma, 2015). Next, instructors must develop the questions based on their approach and inquiry goals. Given the formative nature of the feedback, questions may need to be adapted on an ad hoc basis, and those questions should have various qualitative, dialogic qualities (Harvey, 2003) including open-endedness, specificity, and covering a broad base of issues (see Table 3 for more details).

Instructor

Instructors act in the Instructor role of the Life Cycle of Student feedback by interacting with students to administer the tool and with their feedback. The three main Instructor themes are ease of use, minimal questions at a time, and continuous administration. Ease of use and brevity are crucial affordances, especially for instructors who repeatedly inquire of their students. Ease enables the "data collected . . . to be integrated into a regular cycle of analysis, reporting, action and feedback" (Harvey, 2003). Stieger and Burger (2010) noted that minimal questions should be asked because of the need for quick turnaround with formative feedback. Among the 10 studies in the tool review (Table 6), the number of minimal questions at a time varied widely (see Table 4). Half of the articles exhibited ad hoc or weekly cadences of a single question or open comment form, while Stieger and Berger (2010) administered up to 5 questions, and Winchester and Winchester (2014) asked, an unusually high, 14 questions on a weekly basis.

Lastly, instructors interact with students through feedback. The literature recommended various properties of formative student feedback, including a need for student anonymity, timeliness of instructor response, and bidirectional feedback exchanges. Svinicki (2001) noted that student anonymity is important, because it mollifies the student fear that instructors would retaliate if they knew who gave negative comments or suggestions, thus paving the way for "open"-ended comments. Goldfarb and Morrison (2014), Hendry (2001), and Wagner (2015) particularly underscored timeliness of feedback collection and response, because they are key to student guidance. Timeliness also helps make feedback a better measure by avoiding primacy (Stieger & Burger, 2010) and recency (Steiner & Rain, 1989) effects that result from emotional carryover when self-reported data collection occurs at a different time than episodes in question. Lastly, two-way communication in student feedback is important because it validates students, helping them see that their feedback is going somewhere (Ravelli, 2000). Even speaking from the days before significant technology mediation of feedback, Mosteller (1989) emphasized that instructors should respond to as many requests as possible.

Table 4

Formative Feedback Cadence and Question Frequencies

Study	Tool	Cadence	Number of Questions
(Desai, 2014)	Twitter	Ad hoc, optional	None, 1 open-ended invitation to give feedback
(Hendry et al., 2001)	Survey Link	Ad hoc, optional (1 per web page)	None, 1 open-ended invitation to give feedback
(Wagner et al., 2015)	Survey link	Ad hoc, optional	1 Open-ended
(Hendry et al., 2001)	Survey Link	Weekly Group Survey	None, 1 open-ended invitation to give feedback
(Goldfarb & Morrison, 2014)	Email	Weekly	None, 1 open-ended invitation to give feedback
(Chen & Chen, 2012)	Twitter	Weekly	3 Open-ended
(Stieger & Burger, 2010)	Twitter	Weekly	5: Face-to-face, 3 closed- and 2 open-ended 3: Online, 2 closed- and 1 open-ended
(Winchester & Winchester, 2014)	LMS Survey Link	Weekly	14 closed-ended
(Henrie, Bodily, Manwaring, & Graham, 2015)	DropThought	9 times: 1 per assignment	2: 1 overall rating, and 1 open-ended feedback prompt
(Foth et al., 2012)	Reframe	Twice per class period	5: 2 close-ended before class and 2 after class, plus 1 optional open-ended comment area)

Evaluator

Instructors take on the Evaluator role by analyzing the instructor-prompted student feedback. Feedback analysis is arguably the most time consuming and burdensome step in the formative feedback process (Harvey, 2003; Knol, 2013; Winchester & Winchester, 2011a). Surprisingly, little description and few suggestions were provided in the analysis steps of student feedback beyond instructors reviewing scores and reading and responding to comments. For example, Chen and Chen (2012) noted that required weekly reflections increased the instructor workload significantly, and hence they recommended optional reflection in order to reduce workload burden. Winchester and Winchester's (2011a) study was the only exception to the dearth of feedback analysis descriptions, which delineated instructors' superficial and in depth reflection practices that resulted from feedback. For example, instructors who reflected superficially were reactive and hesitated action, because they felt there was not enough feedback to act on. Whereas instructors who reflected proactively used the feedback as an opportunity to question how things were done (Winchester & Winchester, 2011a). Lastly, as in any evaluation, data needs to be summarized and aggregated. Given the formative feedback context, however, little time remains after collection to make decisions regarding what parts of the course or instructional practices to stop, keep, or start doing (DeLong, n.d.). Foth et al. (2012) reported that the 2 instructors who used the Reframe app each class period to collect student feedback found, "The manual processing meant that trend data was not available, and the information couldn't be shared quickly enough with students, which reduced one of the core benefits: immediacy of feedback" (p. 153). Unfortunately, none of the reviewed articles followed Stieger and Burger's (2010) recommendation for a quick analysis of student feedback.

Interventionist

Instructors act as Interventionists by following up with the students on their feedback. Foth et al. (2012) claimed that the intervention's success depends on how it is shared with students. Two forms of sharing the feedback with students were observed in the reviewed manuscripts. First, some instructors placed importance on reviewing feedback before the next session of class (Foth et al., 2012; Mosteller, 1989), whereas Goldfarb and Morrison (2014) emphasized the importance of periodic summaries of the anonymized raw feedback. In both cases, it was important to demonstrate to the students that they had been heard. For instance, Ravelli (2000) found that student/instructor interaction increased when instructors opened discussion on their own teaching. Ravelli (2000) also found that students enjoyed giving feedback during the school term more than after the fact, because it benefited themselves more and they liked taking part in the teaching process.

Acting on the feedback is idiosyncratic to the situation and instructor. Goldfarb and Morrison (2014) demonstrated, however, the importance of not leaving follow up actions to chance, but having an established follow up structure. An obvious, yet important, practice of sharing what actions the instructor took as a result of the student feedback provides the students with a sense of ownership in their learning (Aultman, 2006) and that their feedback is valued (Crews & Curtis, 2011; Hendry et al., 2001; Svinicki, 2001). Lastly, taking action based on formative student feedback is important because it only results in benefit to the students who gave the feedback when they trust the feedback system. Wagner et al. (2015) found, for example, when students trusted the feedback system they gave positive evaluations of instructors and a more frequently expressed desire to continue in the field of study.

The amalgamation of instructional feedback practices from the formative feedback literature has been organized into themes that together illustrate the Life Cycle of Student Feedback. Instructors capture feedback regarding activities and content from specific occasions, so that it has relevance to the students giving the feedback, as well as to themselves for making sense of it. Instructors taking action and following up with students completes the circle, where students feel valued for having been asked for their input and by seeing positive changes as a result. Formative feedback tools must support this lifecycle.

The Growth of Formative Student Feedback and Tools

Several calls have been made for the use of formative feedback from students for the improvement of instruction (Beaty, 1997; Dennen & Bonk, 2007; Hendry et al., 2001; Jahangiri, Mucciolo, Choi, & Spielman, 2008; Jara & Mellar, 2010; Lewis, 2001; Wagner et al., 2015; Woloschuk, Coderre, Wright, & McLaughlin, 2011). As a developing area of research, formative feedback from students has its roots in formative evaluation, a term coined by Scriven (1967), but as Cambre (1981) noted, it has been a long-standing, informal practice by educators (Weston et al., 1995). Action research in teaching (Kember & Kelly, 1992) has also provided a framework for formative collection of student feedback.

Nevertheless, formative student feedback has not been a widespread practice formally. For example, as Stieger and Burger (2010) noted from a study on feedback use in U.S. and Canadian medical schools (Abrahams & Friedman, 1996), only 4 of 79 reporting schools captured student feedback on a weekly basis. As noted before, Stieger and Buger (2010) suggested that formative feedback has been too cumbersome for widespread adoption, citing high administrative work, displacement of in-class time, and high volumes of data to analyze and act upon. This view is supported by studies reviewed herein (Chen & Chen, 2012; Foth et al., 2012; Knol, 2013; Winchester & Winchester, 2011a). It appears the lack of adoption of formative student feedback has not been for lack of utility or benefit, but for the missing affordances in the tools.

Since the turn of the century, most institutions have adopted online survey tools to replace in-class, paper surveys (Gamliel & Davidovitz, 2005; Hmieleski & Champagne, 2000; Nulty, 2008). Several assertions have been made regarding the potential that online, end-of-course surveys have over in-class, paper surveys (Capa-Aydin, 2014; Crews & Curtis, 2011;

Nevo, McClean, & Nevo, 2010; Ravelli, 2000) including overcoming administration irregularities (Ory, 1990; Simpson & Siguaw, 2000), administration costs (Ballantyne, 2003; Bothell & Henderson, 2003; Morrison, 2013) and in-class time pressure (Ballantyne, 2003; Tucker, Jones, Straker, & Cole, 2003). Though online surveys provide more rapid feedback than paper-based approaches (Anderson, Cain, & Bird, 2005; Spooren, Brockx, & Mortelmans, 2013), student evaluations of instruction are administered largely at the end-of-term (Gravestock & Gregor-Greenleaf, 2008).

More recently, however, shifts in technology have made formative student feedback collection more plausible. The recent proliferation of mobile devices such as tablets and smartphones has ushered in a wide variety of highly downloaded apps that facilitate widespread communication among college students and the public more broadly. Several apps for rating experiences have emerged, making real-time experience rating commonplace. Yelp!, for example, saw more than 66 million downloads as of April 2015 ("Xyo - Apps to the people," n.d.). Higher education has its own version, the often-vilified (Jarvis, 2009; Potter, 2008) ratemyprofessors.com, which is the most frequently used external rating site in education, but known for suspect ratings (Spooren et al., 2013). Other apps common for social networking such as Snapchat, Facebook, Twitter, and Instagram have made communication through mobile devices ubiquitous in higher education today (Chen & deNoyelles, 2013). Consequently, the ways in and the frequency by which students can be reached for feedback on their learning experiences have grown to a level many would have not imagined even a decade ago.

At first glance, the recent proliferation of communication technology makes it seem surprising that formative student feedback has not seen more usage. A closer look at the common qualitative approach by early formative feedback practitioners, however, may explain why few educators take on such a practice. Early examples of formative student feedback relied on paper-based surveys in class, like Mosteller's (1989) "muddiest point in the lecture" questions or Angelo and Cross's (1993) minute papers where instructors at the end of or throughout class would ask students to reflect on or sum up what stood out most from the lecture. All those student answers would have to have been read and synthesized after each session. Though effective at capturing meaningful feedback, such approaches have been loath to widespread adoption.

The overhead required for students and instructors to manage paper-based formative feedback, particularly in large gateway classes, relegates such a practice to instructors only willing to submit to herculean efforts. Clearly, the widespread adoption of formative feedback practices is largely dependent on the technology available to instructors and students to manage the process. Hence, a closer look at the tools that support formative feedback is merited, particularly as they relate to the Lifecycle of Student Feedback.

Analysis of Current Formative Student Feedback Tool Affordances

Tool affordance has a significant impact on quality and quantity of formative student feedback (Stieger & Burger, 2010). As researchers have experimented with formative feedback tools, a variety of recommendations have emerged, yet more from their individual experiences than from the literature. In order to review the tools used in formative feedback practice, we created a set of questions (see Table 5) that correspond to the recommended properties and process steps of formative student feedback as identified in the Lifecycle of Student Feedback framework. Table 5 shows the tool affordance of interest for each question. In this section, we answer, *how do the reviewed formative feedback tools support formative feedback practice?*

Table 5

Formative Feedback Tool Questions

Feature	Question	Affordance
Mode	What type of tool is it?	N.A.
Commercial Availability	Is the tool commercially available?	Enables widespread use and replicability
Purpose	Does the tool scaffold the defining of why instructors are asking for feedback?	Grounds questions in a sense-making framework
Editable Questions	Can instructors modify questions, create new questions over time?	Enables adaptation to needs of a particular point in time
Cadence	How frequent are the reported feedback solicitations?	Situates feedback within specific occasions of instructional and learning practice
Student Ratings	Does the interface feature a rating scale?	Situates comment within and overall feeling
Student Comments	Does the interface feature an open-ended comment form?	Enables students to express the specifics, range, depth, and polarity of their experience
Visibility of Feedback	What level of visibility to others does the student submitted feedback have?	Communicates to students that their comments are being heard
Anonymity	Is the student feedback anonymous?	Enables trust in the feedback system
Bi-directionality	Can instructors reply to student feedback submission?	Enables help requests
Output data	How are the feedback results presented to the instructor?	Determines ease of decision making
Result Sharing	Does the tool facilitate sharing with the students the impact of their feedback?	Communicates to students that they are valued contributors
Instructor Follow up	Does the tool facilitate tracking follow up actions?	Enables consistent feedback practice

The results of the questions listed in Table 5 are summarized in Table 6. The term affordances comes from Gibson's (1979) and Gaver's (1991, 1996) work to describe affordances as characteristics of things in the world that signal what can be done with them, such as how a doorknob with a ridge can be pushed or pulled (Cook & Brown, 1999), and here refer to the aspects of the tools that enable instructors to take action in their practice in the real world.

Commercial Availability

Twitter, DropThought, survey tools, and email are commercially available, whereas Reframe (2012) and Wagner's (2015) abuse reporting system were both developed in-house and are not commercially available. The feedback system in Goldfarb and Morrison (2014), however, featured email for the direct collection of feedback, but the overall system for managing the data in the emails was manual and not commercially available. The commercially available tools support widespread use, though they were not specifically built to support formative feedback practice. Whereas the in-house built tools were designed for specific, formative feedback use cases (see below) even though they are not available for other researchers or instructors to use. What is missing is a commercially available tool built specifically to support the Lifecycle of Student Feedback.

Purpose

None of the tools reviewed supported the critical step of defining a purpose for the feedback inquiry, or the theory of teaching from which the questions come. Consequently, it would be hard to imagine how this practice would be facilitated in any consistent way as formative student feedback grows in usage, thus being in danger of only identifying low-hanging-fruit rather than meaningful change (Biggs & Tang, 2011; Bowden & Marton, 2003).

The purposes of the feedback collection in the reviewed studies, however, were declared. In the mobile apps, all three of the reviewed articles that featured Twitter had instructors that used it for formative evaluation of the class and instruction. Two of the studies (Chen & Chen, 2012; Stieger & Burger, 2010) were for classroom-based classes. The other (Desai, 2014) was for evaluation of clinical rotations by medical students. The article that featured DropThought investigated which had more impact on student satisfaction in blended learning, assignment mode or assignment design (Henrie et al., 2015).

For Reframe, the in-house-developed mobile app by Queensland University of Technology (Foth et al., 2012), the focus was on student check-in mood and student preparedness at the beginning of class, and overall session feedback at the end of class. Wagner et al. (2015) reported on a mobile web page that medical students in clinical rotations used to report anonymously on abuse of students by residents. In the survey group of tools (Hendry et al., 2001; Ravelli, 2000; Winchester & Winchester, 2011b), two featured course evaluation and the other, teacher evaluation. The one email-based approach (Goldfarb & Morrison, 2014) focused on medical curriculum reform as the unit of analysis.

Table 6

Analysis of Formative Feedback (Student to Instructor) Tool Affordances in Published Literature

Tool	Twitter	Twitter	Twitter	Reframe	DropThought*	Single Q.	Survey	Survey	Survey	Email
1st Author	Chen	Desai	Stieger	Foth	Henrie	Wagner	Winchester	Hendry	Ravelli	Goldfarb
Mode	Mobile App	Mobile App	Mobile App	Mobile App	LMS Form	Mobile Web	LMS Link	Website Link	Email Link	Email
Purpose	Course Eval.	Clinical Rotation Eval.	Course Eval.	Course Eval.	Blended Learning Engagement	Abuse Report	Teacher Eval.	Course Eval.	Course Eval.	Curricu-lum Reform
Edited Qs	Yes	Yes	Yes	No	Yes	No	No	No	No	No
Cadence	Weekly	Ad Hoc	Weekly	Per Class	Per assign.	Ad Hoc	Weekly	Ad Hoc	Ad Hoc	Per Class
Rating	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes
Comment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bidirectional	Yes	Yes	Yes	No	Yes	No	No	No	No	No
Anonymity	Pseudo	No	Pseudo	Yes	Yes	Yes	Yes	No	Yes	Yes
Visibility	Group	Group	Instructor	Instructor	Instructor	Cohort	Instructor	Instructor	Instructor	Instructor
Follow Up	No	No	No	No	No	No	No	No	No	No
Result Share	No	No	No	No	No	Yes	No	Annually	No	Yes
Data Format	Raw Data	Raw Data	Raw Data	Raw Data	Dashboard / Raw Data	Raw Data	Report	Emails	Report	Raw Data
Commercial	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes

* The first author designed DropThought as a commercially available product. Its core platform is free, though DropThought generates revenue from institutions through add-ons to the platform. The first author was employed by DropThought during part of the writing of this article.

Though some of the tools embodied the purpose imbued by the institution (abuse reporting and curriculum reform), most of them focused on data gathering that instructors carried out individually. None of these tools scaffolded the inquiry purpose process, as seen in tools that facilitate instructor reflection through video analysis (Rich & Hannafin, 2009). The lack of procedural reification at the instructor-level may be fine if the purpose is clearly supported by the institution, but this leaves at chance whether the tool affords theory-based instructional improvement for individual instructors, particularly regarding their theory of teaching (Biggs & Tang, 2011) or personal goals (Fuhrmann & Grasha, 1983). Without purposeful inquiry and design for the improvement of instruction, formative student feedback could fall prey to the rote review of feedback like seen with end-of-course surveys.

Editable Questions

Can instructors modify questions, create new questions over time? The research from the Twitter articles (Chen & Chen, 2012; Desai, 2014; Stieger & Burger, 2010) could have called for the instructors to vary questions over time, though this was not the case. Both Chen and Chen (2012) and Stieger and Burger (2010) used fixed questions throughout the school term, while Desai (2014) reported that participants used Twitter to provide informal feedback after each clinical round, rather than respond to specific questions.

Neither the Reframe app (Foth et al., 2012) nor the mobile web form reported by Wagner et al. (2015) allowed for instructor-prompted questions; the former provided a generic form and the latter had a pre-written question asking for abuse reports. The articles featuring surveys (Hendry et al., 2001; Ravelli, 2000; Winchester & Winchester, 2011b) and email feedback (Goldfarb & Morrison, 2014) modes did not report changing questions, though both technologies could have facilitated question updating over time. Henrie et al. (2015), on the other hand, used

DropThought to add questions on assignments throughout the school term (Henrie, Personal Conversation, n.d.).

Despite the strong calls for question updating in formative feedback tools (Aultman, 2006; Lewis, 2001; Mosteller, 1989; Ravelli, 2000) only one of the studies used the practice, even though eight of the articles included technologies that afforded the possibility. Chen and Chen (2012) concluded however, that varied questions over time would have yielded better quality feedback than was received.

Cadence

What frequency of feedback does the tool support? Twitter affords nearly continuous questions, though 2 of the trio of articles reported weekly feedback in classes (Chen & Chen, 2012; Stieger & Burger, 2010) and the last per clinical rotation at a minimum with ongoing ad hoc feedback in medical education (Desai, 2014). The Reframe app (Foth et al., 2012) also was used every class session while the abuse report mobile page was completely ad hoc (Wagner et al., 2015). The survey articles also split between weekly (Winchester & Winchester, 2011b) and ad hoc (Hendry et al., 2001; Ravelli, 2000), whereas the email-based curriculum reform was per class for feedback capture (Goldfarb & Morrison, 2014). DropThought prompted students for feedback on each blended course assignment, whether face-to-face or online (Henrie et al., 2015). All the tools closely followed the recommendations from the literature to engage in regular analysis cycles (Harvey, 2003) with feedback on a frequent (Tulgan, 1999) and ongoing basis (Fuhrmann & Grasha, 1983; Goldfarb & Morrison, 2014; McKone, 1999).

Frequent and regular feedback from students is an important advantage over summative measures traditionally captured through end-of-course surveys. As Stieger and Burger (2010) noted, summative and formative feedback measure different constructs, the former is more conducive to improvement of instruction.

Formative feedback timing may also diffuse particular biases associated with traditional end-of-course surveys, including primacy (Steiner & Rain, 1989), the overshadowing of initial impression, recency (Dickey & Pearson, 2005), the overshadowing of recent peak experiences, and order effects, such as selection bias (Estelami, 2015) of early vs. late responders.

Student Ratings

Does the interface feature a rating scale? Though qualitative feedback is generally thought to afford more utility for instructional improvement than ratings (Hendry et al., 2001), 5 of the 10 articles reviewed featured tools with rating scales.

Though Twitter does not have a rating scale as part of its interface, Stieger and Burger (2010) arranged a way through the text of the tweets for the students to answer rating scale questions. The other Twitter-based articles did not collect rating data. In Reframe (Foth et al., 2012), rating scales were prominently featured in order to afford social sharing of mood and preparedness for class, as well as a rating on individual course sessions. DropThought features a smiley face rating scale for overall experience on each comment collected (Henrie et al., 2015). Though online survey technology provides rating scale questions, only one (Ravelli, 2000) collected ratings as part of the formative feedback. The email-based curriculum reform (Goldfarb & Morrison, 2014) had appointed students to send email polls to their class peers for each class.

Though summative student feedback affords (Wilson & Ryan, 2012) rote review by instructors with little change in instruction, which may be due to the rating-scale dominated format (Boerboom et al., 2015), the rating scales in formative feedback seem to afford an indexing of experiences and a prioritization of what to reflect on or mark for further discussion, as seen in Goldfarb and Morrison's (2014) report on curriculum reform.

Student Comments

All of the tools reviewed featured a comment area, though it is interesting that Reframe, the in-house-developed mobile app characterized the open form as optional (Foth et al., 2012) rather than the focus of the feedback gathering as in the others. Overall, the tools followed closely with the calls in the literature for feedback to be qualitative (Jara & Mellar, 2010; Lewis, 2001; Ravelli, 2000; Watson, 2003), open-ended (Hendry et al., 2001), and broad-based (Fuhrmann & Grasha, 1983; Potter, 2008).

Visibility of Feedback

Does the tool facilitate sharing of feedback results with the students? Twitter facilitates the sharing of student feedback with the class though it can be kept private between instructor and student as well. In only one of the Twitter studies (Stieger & Burger, 2010) was individual student feedback not shared with the broader class. The Reframe mobile app (Foth et al., 2012) showed class check-in ratings in real-time to the class while the end-of class ratings and comments went to the instructor only. The abuse reporting mobile web system (Wagner et al., 2015) enabled the wider cohort of students across rotations to view anonymized feedback on a periodic basis. The survey studies (Hendry et al., 2001; Ravelli, 2000; Winchester & Winchester, 2011b) and the email-based curriculum reform (Goldfarb & Morrison, 2014) all directed feedback to the instructor. Henrie et al. (2015) did not have the instructors display DropThought feedback to students.

Despite the calls for individual, formative feedbacks to be shared with the broader class (Foth et al., 2012; Mosteller, 1989; Wagner et al., 2015), 6 of the 10 articles described feedback only going to the instructor. Furthermore, none of the tools had an effective way for the feedback to be collected privately and then shared anonymously as needed, a practice called for in the literature. Although it may be recommended to share the student feedback to the broader class, it seems it would take an uncommon level of bravery on the instructor's part to do so. Berk (2005) summed up the feelings many instructors have about student feedback succinctly by channeling Psycho's shower scene, "Why not just whack me now, rather than wait to see those student ratings again." (p. 49).

Student Anonymity

Is the student feedback anonymous? In 8 of the 10 articles, the tools featured anonymous student feedback practice. Only Desai's Twitter feedback for clinical rotations and Hendry's student-centered medical course evaluations used identified feedback. The other 2 Twitter articles (Chen & Chen, 2012; Stieger & Burger, 2010) used pseudonyms as Twitter handles.

Hendry (2001) noted that the students in the program did not object to signed evaluation, but Hendry's institution seems to miss the point about anonymity. Not only is signed feedback biased significantly to be positive more than anonymous feedback (Marsh, 1984), but the subject matter of feedback changes when students are in fear of reprisal (Svinicki, 2001).

Wagner (2015) also found from surveys that when students trusted the anonymity of the formative reporting system they were 5 times less likely to perceive abuse from the residences in charge of their rotations. Anonymity of the tool and students' trust in its anonymity play a significant factor in the quality of data the tools afford. These results support calls in the literature for student feedback to be anonymous.

Bi-directionality

Among the tools reviewed, only Twitter and DropThought facilitated two-way interaction. The other mobile and survey tools did not afford bi-directionality. Email is generally two-way, but as Goldfarb and Morrison (2014) described, selected students in charge of collecting feedback sent feedback requests to their peers and then passed on anonymized messages to administrators and faculty. There was not a way for individual student feedback messages to be replied to directly. The feedback tool affordances largely did not meet the criteria discussed in the literature as being important for formative feedback. As Mostellar (1989) argued, it is important to respond to as many requests as possible.

Output Data

A common thread among the tools examined in this review and among the broader literature is the labor intensiveness of qualitative data summarization and sense making (Foth et al., 2012; Knol, 2013; Winchester & Winchester, 2011b). None of the tools afforded summarization or sense making of qualitative data except for DropThought, which had a dashboard that instructors used to manage feedback for their classes, but was not used by the researchers for the analysis of blended learning (Henrie, Personal Conversation, n.d.). Elaborate analysis processes, however, were set up to engage with the resulting data, but outside the collection tools (Desai, 2014; Goldfarb & Morrison, 2014; Hendry et al., 2001; Ravelli, 2000; Wagner et al., 2015; Winchester & Winchester, 2011b). It appears that formative feedback tools overall require significant advances in order to meet Stieger and Burger's (2010) call for quick and easy results.

Result Sharing

Only Twitter in this review facilitates the sharing of changes made in the class as a result of the feedback. This is surprising seeing that students' willingness to give feedback is related to observing changes made or believing that they can be made as a result of their feedback (Crews & Curtis, 2011; Hendry et al., 2001; Svinicki, 2001; Watson, 2003).

Though the tools largely did not afford this important practice, workarounds were created in order to facilitate this communication. Ravelli (2000), for example, reported that students felt more involved in the teaching and learning process. Goldfarb and Morrison (2014), even though email was used to collect student feedback, found that Twitter worked best to notify students of updates to the curriculum as a result of their feedback.

Instructor Follow Up

As Goldfarb and Morrison (2014) suggested, student feedback should have a follow up structure. None of the tools reviewed however, afforded itemization and tracking of changes to be made for ones that have been completed. Various articles cited changes they would make or had made because of the feedback. Some had elaborate structures involving administrators and students (Goldfarb & Morrison, 2014; Wagner et al., 2015).

Supporting Formative Feedback Roles through Tool Design

In this article, the formative feedback literature was reviewed to identify the larger set of practices the instructor carries out when asking for, facilitating the capture of, making sense of, and acting on student feedback, which has been named the Lifecycle of Student Feedback. The identified practices also provided a rubric by which to evaluate how well the tools used in the literature met the recommendations of the same. Though some practices in the Lifecycle of Student Feedback were found to be supported by the technology, such as the writing of qualitative questions and frequent question administration, the current tools also had gaps regarding key practices; hence the importance of investigating *What affordances should be included in effective formative feedback support tools?*

Supporting the Inquirer

In terms of formative feedback, instructors must decide what questions to ask students, and construct the questions. The tools reviewed do not scaffold the focus of inquiry. Processes such as development of a theory of instruction, determining instructor goals for a course, or just deciding what to ask, all happened outside tools used to capture formative feedback. Similarly, in the practice of developing formative feedback questions, the tools did not directly facilitate much of the recommended practice. For example, all of the tools afforded qualitative data collection and most of them afforded instructors writing their own questions, but not necessarily broad-based, open-ended questions. A significant contribution to the formative feedback field would be made if tool developers reified the steps for theory development and question writing through some sort of wizard. To the extent that formative feedback tools can make the purpose of instructional improvement inquiry and the writing of its questions obvious and easy, the wider the impact will be.

Supporting the Instructor

Tool administration for feedback capture and student interaction through feedback, on the other hand appeared to be quick and easy (Stieger & Burger, 2010) across the tools except for Goldfarb and Morrisen's (2014) email-based curriculum reform. Continuous administration was also facilitated by the tools, though more easily by the mobile than the digital tools. Reframe, for example, prompted feedback at the beginning and end of each course session (Foth et al., 2012).

In most cases the tools afforded timely interaction around the feedback, principally because feedback capture was close to the reported experiences.

Though it was possible to carry out formative student feedback with the tools reviewed, the consistency in which the recommendations for formative feedback practice were adhered to was lacking. Anonymous feedback submissions and bi-directionality, for example, were inconsistently exhibited across the tools and their use in the studies. This was in part from the tool affordances. Commercial tools that facilitate a variety of use cases beyond formative student feedback (Twitter, surveys, email) did not support formative feedback best practices in an obvious way. Students had to create anonymous, proxy Twitter accounts with pseudonym handles just used for their course (Chen & Chen, 2012; Stieger & Burger, 2010). Goldfarb and Morrison (2014) reported an elaborate network of students summarizing their peers' email feedback before sending it on to supervisors. Only DropThought (Henrie et al., 2015) provided an out-of-the-box solution to afford both bi-directionality and student anonymity at the same time without any extra steps. Surprisingly, however, even when the tools afforded the recommended practice, some implementations (Desai, 2014; Hendry et al., 2001) did not follow even broadly supported recommendations like keeping student feedback anonymous, as Hendry et al. (2001) reported, "To date students have not expressed any concerns about the lack of confidentiality in this aspect of the system" (p. 330). Hence, given the repeated nature of formative feedback, particularly amidst the many aspects of a course that instructors and students normally engage with, tools must seamlessly support formative feedback best practice consistently.

All the studies in the tool review demonstrated positive outcomes towards the improvement of instruction and courses from formative feedback, including the quick sharing of effective innovations (Desai, 2014; Goldfarb & Morrison, 2014), students engaging in the course improvement process (Aultman 2006; Ravelli, 2000), and increasing student engagement with and likelihood of continuance in the subject (Wagner et al., 2015). These positive developments and others support the preliminary findings (Winchester & Winchester, 2014) that formative feedback can improve instruction over time.

These positive outcomes, however, seem to be related more to what instructors and administrators executed outside the formative feedback tools. Nevertheless, the danger of leaving pivotal processes to chance from one professor to the next, such as effective inquiry and instructor-student interaction, makes it much more likely that meaningful change would not occur across the practice at large.

Supporting the Evaluator

The largest pain points in formative student feedback practice is the combination of volumes of data with the need for quick turnaround and decision making on the part of the instructors. None of the tools reviewed were shown to facilitate an effective student feedback evaluation process, which, as some suggested, defeats a major purpose of having formative feedback (Foth et al., 2012; Stieger & Burger, 2010). Twitter and DropThought both provide interfaces for reviewing student feedback and are searchable for comments, but there was no description of these affordances in the four related articles. Lastly, instructors appear to avoid using formative feedback often because the most difficult parts of the process are not supported by the tools. Hence, formative feedback tool providers should focus on ways of alleviating the analysis and sense making processes of qualitative and quantitative feedback at scale.

Supporting the Interventionist

In the reviewed manuscripts, intervention activities included recognizing student concerns and taking action with the feedback received. In terms of two-way communication within the tools, sharing the feedback received with the class, only Twitter afforded this capability. The remaining studies in the tool review exhibited workaround practices or did not show how they communicated the group's feedback to the students.

With regards to taking action and communicating results to students during the same school term, these practices took place entirely outside of the formative feedback tools. The reviewed tools may have produced the data that enabled continuous improvement of the courses and instruction, but did not facilitate the improvement process itself.

Given that the students' trust in the feedback system has significant impacts on their course experience and how the students feel valued, following up with students regarding the feedback they provide should not be left to chance. Formative feedback tool providers should facilitate feedback sharing and follow up practices observed in the literature in order to have the greatest effect.

Implications

How does one improve on instructional decisions when most of them are tacit and made in real-time? Having an instructor just review course evaluation results at the end-of-term sheds very little light on which decisions lead to those results. Reflection can be an important tool, but it only has the power to improve instruction when instigated by real-time feedback close to the occasions in question (Winchester & Winchester, 2014). With the advancement of technology, the unit of analysis regarding course and instructional improvement has begun to move from the course level to more frequent intervals throughout the course. As such, the understanding of student feedback to instructors must be increased beyond how students are satisfied with the course overall or the validity and reliability of data from summative measures. An examination of how student feedback affects formative decision making on course design and instruction is required. Furthermore, given that formative approaches to student feedback require significant technology mediation in order to be an easy and widespread practice, how tools facilitate formative feedback practice is also critical.

Formative student feedback and the tools that support it are in their infancy. Nevertheless, the variety of tools reviewed all provided data at a cadence that made it possible for instructors to evaluate the effectiveness of the course or learning experience during its enactment. Given the findings of Winchester and Winchester's (2014) longitudinal study that instructor ratings went up by approximately 10% over 2 years when feedback was collected and reviewed on a weekly basis, such a capability has the potential to make a significant impact on course design and instruction. The pain points in the evaluation of student feedback workload recognized across some of the studies (Chen & Chen, 2012; Foth et al., 2012; Stieger & Burger, 2010; Wagner et al., 2015) underscores, however, that formative feedback tool design should examine the recommended steps of the practice and properties of the feedback.

Educational technologists, instructors and administrators should take note of the Lifecycle of Student Feedback in the design and selection of tools that support the effective feedback practices outlined in the literature. Apart from anonymous and frequent feedback capture, most of the formative feedback practice responsible for improvement of instruction resides outside the formative feedback tools. For example, the tracking and reporting of improvements in curriculum and instruction, though highly desired by students, appears to be infrequently practiced and has seen little attention in the literature and tool development. More should be done to ensure innovations and updates to courses to show a clear link between the changes made, their feedback, and the resulting impacts, thus centering the design of instruction and curriculum on the student.

Lastly, in order for tools to consistently facilitate instructional improvement they must embody the practices that lead to it. Better standardization and facilitation of inquiry goals, sensemaking, and feedback follow up would add considerable ease to the managing of student feedback practice. Without a standard of practice for formative improvement of instruction and a tool that supports those practices, instructional improvement may be left to only the most ambitious instructional programs.

References

Abrahams, M. B., & Friedman, C. P. (1996). Preclinical course-evaluation methods at US and Canadian medical schools. *Academic Medicine*, *71*(4), 371–4.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*(2), 153–166.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441.

Anderson, H. M., Cain, J., & Bird, E. (2005). Online student course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education*, *69*(1–5), 34–43.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2 edition). San Francisco, CA: Jossey-Bass.

Annan, S. L., Tratnack, S., Rubenstein, C., Metzler-Sawin, E., & Hulton, L. (2013). An integrative review of student evaluations of teaching: Implications for evaluation of nursing faculty. *Journal of Professional Nursing*, *29*(5), e10–e24.

Aultman, L. P. (2006). An unexpected benefit of formative student evaluations. *College Teaching*, *54*(3), 251–285.

Ballantyne, C. (2003). Online evaluations of teaching: An examination of current practice and considerations for the future. *New Directions for Teaching and Learning*, *2003*(96), 103–112.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, *87*(4), 656-665.

Beaty, L. (1997). *Developing your teaching through reflective practice*. Birmingham: SEDA.

Beran, T., Violato, C., & Kline, D. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "Consequential Validity" study. *Canadian Journal of Higher Education*, *35*(2), 49–70.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, *17*(1), 48–62.

Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education*, *41*(3), 221–238.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (3rd edition). Poland: McGraw-Hill Education (UK), Open University Press.

Boerboom, T. B. B., Stalmeijer, R. E., Dolmans, D. H. J. M., & Jaarsma, D. A. D. C. (2015). How feedback can foster professional growth of teachers in the clinical workplace: A review of the literature. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2015.02.001

Bothell, T. W., & Henderson, T. (2003). Do online ratings of instruction make $ense? *New Directions for Teaching and Learning*, *2003*(96), 69–79.

Bowden, J., & Marton, F. (2003). *The university of learning*. London: Psychology Press.

Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance. Jossey-Bass Higher and Adult Education Series.* ERIC. Retrieved from http://eric.ed.gov/?id=ED368305

Cambre, M. A. (1981). Historical overview of formative evaluation of instructional media products. *ECTJ*, *29*(1), 3–25. https://doi.org/10.1007/BF02765189

Capa-Aydin, Y. (2014). Student evaluation of instruction: comparison between in-class and online methods. *Assessment & Evaluation in Higher Education*, *0*(0), 1–15. https://doi.org/10.1080/02602938.2014.987106

Carle, A. C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers & Education*, *53*(2), 429–435. https://doi.org/10.1016/j.compedu.2009.03.001

Cashin, W. E. (1999). Student ratings of professors: Uses and misuses. In P. Seldin & P. Hutchings (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 25–44). Bolton, MA: Jossey-Bass.

Chen, B. & deNoyelles, A. (2013). *Exploring students' mobile learning practices in higher education*. EDUCAUSE Review. Retrieved from http://www.educause.edu/ero/article/exploring-students-mobile-learning-practices-higher-education

Chen, L., & Chen, T.-L. (2012). Use of Twitter for formative evaluation: Reflections on trainer and trainees' experiences. *British Journal of Educational Technology*, *43*(2), E49–E52.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16–30.

Cook, S. D., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*(4), 381–400.

Crews, T. B., & Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education*, *36*(7), 865–878.

DeLong, T. J. (n.d.). Three questions for effective feedback. Retrieved July 29, 2015, from https://hbr.org/2011/08/three-questions-for-effective-feedback

Dennen, V. P., & Bonk, C. J. (2007). We'll leave the light on for you: Keeping learners motivated in online courses. In *Online and distance learning: Concepts, methodologies, tools, and applications*. L.A. Tomei, (Ed.). (Vol. 2, pp. 704–715). IGI Global.

Desai, B. (2014). A novel use of Twitter to provide feedback and evaluations. *The Clinical Teacher*, *11*(2), 141–145.

Dickey, D., & Pearson, C. (2005). Recency effect in college student course evaluations. *Practical Assessment, Research & Evaluation*, *10*(6). Retrieved from http://pareonline.net/pdf/v10n6.pdf

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, *11*(1), 37–46.

Estelami, H. (2015). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing Education*, *37*(1), 54–64.

Fabris, C. (2015, February 4). Anonymous feedback, fine. Insults? Not on these platforms. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Anonymous-Feedback-Fine/151609/

Foth, M., Fitz-Walter, Z., Ti, J., Russell-Bennett, R., & Kuhn, K.-A. (2012). Please take out your phones: On the spot solicitation of student feedback in class. In *Proceedings of the 24th*

Australian Computer-Human Interaction Conference (pp. 150–153). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2414563

Franklin, J., & Theall, M. (1989). Who reads ratings: Knowledge, attitude, and practice of users of student ratings of instruction. Retrieved from http://eric.ed.gov/?id=ED306241

Fuhrmann, B. S., & Grasha, A. F. (1983). *A practical handbook for college teachers*. Boston, MA: Scott Foresman & Company.

Gamliel, E., & Davidovitz, L. (2005). Online versus traditional teaching evaluation: Mode can matter. *Assessment & Evaluation in Higher Education*, *30*(6), 581–592.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 79–84). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=108856

Gaver, W. W. (1996). Situating action II: Affordances for interaction: The social is material for design. *Ecological Psychology*, *8*(2), 111–129.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Goldfarb, S., & Morrison, G. (2014). Continuous curricular feedback: A formative evaluation approach to curricular improvement. *Academic Medicine February 2014*, *89*(2), 264–269. https://doi.org/10.1097/ACM.0000000000000103

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Higher Education Quality Council of Ontario Toronto. Retrieved from http://www.celt.iastate.edu/pdfs-docs/classclimate/research/2008_gravestock_course_evaluations_research_models_publication.pdf

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1209-1217.

Harvey, L. (2003). Student feedback. *Quality in Higher Education*, *9*(1), 3–20. https://doi.org/10.1080/13538320308164

Hendry, G. D., Cumming, R. G., Lyon, P. M., & Gordon, J. (2001). Student-centred course evaluation in a four-year, problem based medical programme: Issues in collection and management of feedback. *Assessment & Evaluation in Higher Education*, *26*(4), 327–339. https://doi.org/10.1080/02602930120063484

Henrie, C. (n.d.). Personal Communication.

Henrie, C. R., Bodily, R., Manwaring, K. C., & Graham, C. R. (2015). Exploring intensive longitudinal measures of student engagement in blended learning. *The International Review of Research in Open and Distributed Learning*, *16*(3). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/2015

Hmieleski, K., & Champagne, M. V. (2000). Plugging in to course evaluation. *The Technology Source Archives at the University of North Carolina*. Retrieved from http://technologysource.org/article/plugging_in_to_course_evaluation/

Jahangiri, L., Mucciolo, T. W., Choi, M., & Spielman, A. I. (2008). Assessment of teaching effectiveness in U.S. dental schools and the value of triangulation. *Journal of Dental Education*, *72*(6), 707–718.

Jara, M., & Mellar, H. (2010). Quality enhancement for e-learning courses: The role of student feedback. *Computers & Education*, *54*(3), 709–714.

Jarvis, A. M. (2009, November 12). Defending RateMyProfessors.com. Retrieved from http://chronicle.com/blogs/profhacker/defending-ratemyprofessorscom/22828

Kember, D., & Kelly, M. E. (1992). *Improving teaching through action research*. Campbelltown, N.S.W: Higher Education Research and Development Society of Australasia.

Knol, M. H. (2013). Improving university lectures with feedback and consultation. Retrieved from http://dare.uva.nl/en/record/443738

Lang, J. W., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, *35*(3), 187–205.

Lewis, K. G. (2001). Using midsemester student feedback and responding to it. *New Directions for Teaching & Learning*, *2001*(87), 33-44.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*(5), 707-754.

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, *99*(4), 775-790.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187-1197.

McKeachie, W. J., & Kaplan, M. (1996). Persistent problems in evaluating college teaching. *AAHE BULLETIN*, *48*, 5–8.

McKone, K. E. (1999). Analysis of student feedback improves instructor effectiveness. *Journal of Management Education*, *23*(4), 396–415.

Morrison, K. (2013). Online and paper evaluations of courses: A literature review and case study. *Educational Research and Evaluation*, *19*(7), 585–604.

Mosteller, F. (1989). The "Muddiest Point in the Lecture" as a feedback device. *On Teaching and Learning: The Journal of the Harvard-Danforth Center*, *3*, 10–21.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, *27*(2), 187–198.

Nevo, D., McClean, R., & Nevo, S. (2010). Harnessing information technology to improve the process of students' evaluations of teaching: an exploration of students' critical success factors of online evaluations. *Journal of Information Systems Education*, *21*(1), 99–109.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301–314.

Ory, J. C. (1990). How can student ratings of instruction be used for assessment? *Assessment Update*, *2*(3), 9–11.

Potter, C. (2008, April 27). How sharper than a serpent's tooth it is...or is it? The radical rates rate my professors.com - tenured radical. Retrieved July 29, 2015, from http://chronicle.com/blognetwork/tenuredradical/2008/04/how-sharper-than-serpents-tooth-it-isor/

Potter, J. (2008). Re-designing an MA module to foster agency, engagement and production in online social software. *Reflecting Education*, *4*(1), 81–91.

Ramsden, P. (2003). *Learning to teach in higher education*. London & New York, NY: Routledge.

Ravelli, B. (2000). Anonymous online teaching assessments: Preliminary findings. Presented at the Annual National Conference of the American Association for Higher Education, Charlotte, NC. Retrieved from http://eric.ed.gov/?id=ED445069

Rich, P. J., & Hannafin, M. (2009). Video annotation tools technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education*, *60*(1), 52–67. https://doi.org/10.1177/0022487108328486

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, *38*(5), 575–592.

Scriven, M. S. (1967). *The methodology of evaluation* (Vol. 1). Chicago, IL: Rand McNally.

Shao, L. P., Anderson, L. P., & Newsome, M. (2007). Evaluating teaching effectiveness: Where we are and where we should be. *Assessment & Evaluation in Higher Education*, *32*(3), 355–371.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, *22*(3), 199–213.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching the state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Steiner, D. D., & Rain, J. S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, *74*(1), 136–142. https://doi.org/10.1037/0021-9010.74.1.136

Stevens, J. J., & Aleamoni, L. M. (1984). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science*, *13*(4), 285–304.

Stieger, S., & Burger, C. (2010). Let's go formative: Continuous student ratings with web 2.0 application Twitter. *Cyberpsychology, Behavior, and Social Networking*, *13*(2), 163–167.

Svinicki, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning*, *2001*(87), 17–24.

Theall, M., & Franklin, J. (1990). Student ratings in the context of complex evaluation systems. *New Directions for Teaching and Learning*, *1990*(43), 17–34.

Tucker, B., Jones, S., Straker, L., & Cole, J. (2003). Course evaluation on the web: Facilitating student and teacher reflection to improve learning. *New Directions for Teaching and Learning*, *2003*(96), 81–93.

Tulgan, B. (1999). Fast feedback. *Employment Relations Today*, *26*(2), 73–83. https://doi.org/10.1002/ert.3910260209

Wagner, J. P., Tillou, A., Nguyen, D. K., Agopian, V. G., Hiatt, J. R., & Chen, D. C. (2015). A real-time mobile web-based module promotes bidirectional feedback and improves evaluations of the surgery clerkship. *The American Journal of Surgery*, *209*(1), 101–106.

Watson, S. (2003). Closing the feedback loop: Ensuring effective action from student feedback. *Tertiary Education & Management*, *9*(2), 145–157.

Weston, C., McAlpine, L., & Bordonaro, T. (1995). A model for understanding formative evaluation in instructional design. *Educational Technology Research and Development*, *43*(3), 29–48. https://doi.org/10.1007/BF02300454

Wilson, J. H., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching*, *62*(1), 20–24.

Wilson, J. H., & Ryan, R. G. (2012). Formative teaching evaluations: Is student input useful? in *effective evaluation of teaching: a guide for faculty and administrators* (pp. 22–29).

Society for the Teaching of Psychology. Retrieved from http://stp.wildapricot.org/Resources/Documents/ebooks/evals2012.pdf#page=26

Winchester, T., & Winchester, M. (2010). Exploring the utility of using a VLE for weekly student evaluations of teaching. In *International Conference on Information Communication Technologies in Education (ICICTE), Corfu, Greece*. Retrieved from http://www.harper-adams.ac.uk/aspire/files/WinchesterTLWTdissertation.pdf

Winchester, T., & Winchester, M. (2011a). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development*, *16*(2), 119–131.

Winchester, & Winchester. (2011b). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development*, *16*(2), 119–131.

Winchester, & Winchester. (2012). If you build it will they come?; Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *37*(6), 671–682.

Winchester, & Winchester. (2014). A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology*, *45*(1), 112–124. https://doi.org/10.1111/bjet.12019

Woloschuk, W., Coderre, S., Wright, B., & McLaughlin, K. (2011). What factors affect students' overall ratings of a course? *Academic Medicine*, *86*(5), 640–643.

Xyo - Apps to the people. (n.d.). Retrieved April 14, 2015, from http://xyo.net/

ARTICLE #2: TOWARDS ACTIONABLE COURSE DESIGN DATA THROUGH REAL-TIME FEEDBACK

Running head: Actionability of Real-Time Student Feedback

Towards Actionable Course Design Data through Real-Time Feedback

Michael A. Atkisson, Ross Larsen, and Richard E. West

Brigham Young University

Abstract

This study proposes a method to capture real-time feedback from students to instructors through electronic feedback prompts situated throughout a course. The student response rate for comments was 87% overall with 70% of students (N=276) responding on 9 out of 9 occasions or more. The high engagement also produced a wide variety of feedback topics (2,092 comments into 8 topic and 27 subtopic categories), which helped identify what comment properties increased the likelihood of comment actionability. We found strong negative correlations (r=0.97 - 0.56, p=0.05) between a comment's actionability and its sentiment level across various topical groupings, including whether the comment was a request, suggestion, question, or statement. Through hierarchical general linear modeling we also found that a student authorship interclass correlation accounted for 17% of the variance in comment actionability (Design Effect=2.8466) and that various topic predictors increased the likelihood of comment actionability, including accessibility (Odds Ratio=9.1432, p= 0.0) and usability issues (Odds Ratio=5.3069, p= 0.0). Lastly, we found the ratio of actionable comments to be higher in end-of-course feedback (0.28, Z=2.9, p=0.0) than in real-time feedback (0.19, Z=2.9, p=0.0), a surprising result that may be due to a priming effect with real-time feedback. The method has practical implications for instructional design, teaching, and interventionist research with students.

Introduction

End-of-course student feedback is widespread in higher education (Berk, 2005), often as a summative evaluation of instruction (Seldin, 1989). With all this feedback data, it would seem useful to formatively improve courses and instruction. Overall, there is little evidence that stakeholders in course quality, (e.g., instructors, instructional staff, instructional designers, academic policy leaders) use student feedback for improvement in any systematic way (Maistre & Weston, 1996; Weston, Le Maistre, Mcalpine, & Bordonaro, 1997). On the other hand, student feedback collected regularly throughout a course, or formative feedback, has shown promise for improving courses and instruction, but there is little research on which properties of formative feedback make it effective feedback. In this study, we examine the properties of formative feedback to understand how comments collected in real-time throughout 15 cohorts of an online professional development course can provide actionable data for course revision decisions.

End-of-Course Surveys

End-of-course survey results are largely used summatively to evaluate the effectiveness of instruction. As Seldin (1989) reported, 88% of liberal arts colleges in the United States (U.S.) use end-of-course surveys for summative decisions. A U.S. Department of Education survey (1991) reported that 91% of 40,000 department chairs from U.S. higher education institutions used end-of-course surveys to rate the effectiveness of instruction (as cited in Berk, 2005).

The widespread practice of students rating instruction at the end-of-course is also hotly debated in education. One side has found the claims regarding instructional effectiveness ratings to be valid (Marsh, 1984; Marsh & Roche, 1997; Spooren, Brockx, & Mortelmans, 2013) and ratings to be reliable measure over time (Marsh, 1984, 2007; Marsh & Roche, 1997). Others have found significant biases in the post-facto ratings about instructor attributes, including minority status of face-to-face instructors (Carle, 2009), perceived attractiveness (Ambady & Rosenthal, 1993), grade leniency (Greenwald & Gillmore, 1997), age (Wilson, Beyer, & Monteiro, 2014), and gender (Basow, 1995). Nevertheless, student feedback through end-of-course surveys is the most common source of data available to instructors about their performance and that of their courses (Abrahams & Friedman, 1996; Berk, 2005; Stieger & Burger, 2010). Though up for debate, post-facto surveys have been argued to provide the most valid evidence of teaching effectiveness (McKeachie, 1997), particularly in near absence of other valid measures (Marsh & Roche, 1997).

Adding to the debate, other researchers have found that end-of course surveys have no long-term impact on the improvement of instruction over time (Lang & Kersting, 2007; Marsh, 2007; Menges & Brinko, 1986; Wilson & Ryan, 2012; Winchester & Winchester, 2014). For example, Marsh (2007) found that among 6,024 undergraduate- and graduate-level university courses taught by 195 instructors over a 13-year period, instruction did not improve or decline over time, (i.e., experience teaching did not result in better teaching ratings). Those who received poor ratings at the beginning of the study continued to receive poor ratings throughout. These findings support other research that shows the rote administration of end-of-course evaluations alone has no effect on the quality of instruction (Menges & Brinko, 1986; Wilson & Ryan, 2012). Furthermore, students' ratings of their instructors' teaching go down on average over time at institutions where instructor ratings were newly implemented, as observed in a 2-year study (Lang & Kersting, 2007) and a 10-year study (Stevens & Aleamoni, 1984). Even though the end-of-course survey ritual alone appears to have had little impact on the improvement of instruction, various studies have shown that through a combination of efforts, including peer review (Berk, 2005) and peer observation (Boerboom, Stalmeijer, Dolmans, & Jaarsma, 2015), the effect can be significant. Nevertheless, these are time-consuming efforts that the vast majority of instructors are unlikely to participate in formally on a regular basis (Thomas, Chie, Abraham, Raj, & Beh, 2014).

Ineffective Use of End-of-Course Feedback

End-of-course surveys' lack of measurable impact on the improvement of courses may be related to a misalignment between instructors who teach a course with little power to change it and the instructional designers and tenured faculty that manage curriculum. Certainly, attentive instructors would be quick to point to changes they have made to their courses based on student feedback. The challenge, however, is that most college course instructors do not have ownership in a course's content or in its design. The Association of Governing Boards of Universities reported (Kezar and Maxey, 2013) that non-tenured track teaching appointments leave instructors without access to curriculum planning and faculty meetings. This is notable considering that the American Association of University Professors (2016) reported that only 29.5% of faculty appointments were tenured or tenure-tracked. Furthermore, for online institutions, many courses are centrally designed by instructional designers rather than the faculty who teach them.

The difficulty with end-of-course survey impact may also be related to 2 problems in the resulting data themselves. First, the data is problematic for decision-making. How course survey data are captured today is largely ratings-focused (Fluit et al., 2010; Gravestock & Gregor-Greenleaf, 2008; Stieger & Burger, 2010), which does not underscore clear actions to take (Biggs & Tang, 2011; Bowden & Marton, 2003; Chapple & Murphy, 1996; Hendry, Cumming, Lyon, & Gordon, 2001; Saffran, Conran, & Lacher, 1994) and is at the expense of qualitative data that would signal what changes to make (Boerboom et al., 2015; Hendry et al., 2001; Woloschuk, Coderre, Wright, & McLaughlin, 2011).

Second, student feedback is almost always collected at the end of the class (Gravestock & Gregor-Greenleaf, 2008; Spooren et al., 2013) for summative decisions about the instructor. At the end of the class, however, is also when students have little incentive to provide meaningful feedback, which results in a greater degree of general comments (Nasser & Fresko, 2002). End-of-course collection is also encumbered by recall biases (Estelami, 2015; Nicol & Macfarlane-Dick, 2006), including the peak-end rule, where the peak emotion from the experience, whether positive or negative, overshadows the average of the individual experiences throughout, especially when the peak emotion is negative (Woloschuk et al., 2011). Together, these factors make it difficult for individual instructors or designers in general to take meaningful course or instructional improvement actions on their own by reviewing student feedback alone. Consequently, only the most concerted efforts by instructors and designers in a large system of support, like through an on-campus center for teaching and learning, are likely to result in significant improvement.

Formative Feedback

If greater and more widespread improvement of instruction and courses is to take place across higher education, better data than what are captured today at the end of courses must identify specific modifications that individual instructors, designers, and policy makers can implement. Several researchers and practitioners have turned to feedback collected from students throughout the school term because it effectively captures more granular feedback (Aultman, 2006; Desai, 2014; Goldfarb & Morrison, 2014; Hendry et al., 2001; McKone, 1999; Ravelli, 2000; Wagner et al., 2015; Winchester & Winchester, 2010, 2011, 2012). Though the term formative would suggest that feedback is being used for improvement (Hickey, 2015), the literature most often differentiates formative feedback from end-of-course surveys by frequency at which feedback is captured, usually at a weekly, per class session, or ad hoc basis throughout the school term (Atkisson & West, n.d.). Nevertheless, recent studies of formative feedback from students to instructors have shown promise in affecting instructor reflection and improvement of instruction (Winchester & Winchester, 2011, 2014), suggesting that the cadence at which feedback is given has a significant impact on instructors' reflection practices and the impact they have on the course (Atkisson & West, n.d.).

For example, Winchester and Winchester (2014) showed in exploratory research with 12 instructors over 2 years that the deeper level of reflection exhibited by instructors in response to weekly formative student feedback, the more increase those instructors received in effectiveness of instruction ratings. These findings support the idea that instructor reflection is more effective throughout a course as opposed to at its end, because instructors would have an easier time recalling situations mentioned in the feedback (Winchester & Winchester, 2014). As Ramsden said, "Evaluation at the end of a course, cannot replace evaluation during it" (2003, p. 242, as cited in Winchester & Winchester, 2012).

Because of the promising benefits of formative feedback for the improvement of student learning experiences (Atkisson & West, n.d.), more research is needed on how to collect, report, and benefit from student feedback for learning design decisions. Furthermore, real-time feedback tools are becoming readily available (Atkisson & West, n.d.). Such tools open the door for the widespread capture of targeted feedback throughout students' academic and social journeys in education, hence the need to understand the nature of feedback giving and capture is at hand. In particular, the general assertion in formative feedback literature that its effectiveness is due to its granularity insufficiently explains how it is more useful than end-of-course survey data for course and instructional design decisions in practice. We not only want to establish a reliable method for capturing actionable student feedback, but also to explain why some types of feedback may be more effective than others.

From Real-Time to Actionable Feedback

The notion of real-time feedback is not an increased cadence of long-form, end-of-course surveys, but is delivered through a short, or even single question, electronic feedback that can easily be place in the flow of activity (Atkisson & West, n.d.). In this study, we anticipated that real-time feedback would more likely be actionable for course design decisions than comments made at the end-of-course. Actionability colloquially describes the ease that data afford decision makers to act. We view actionability in terms of dynamic affordance (Cook & Brown, 1999), which suggests that student feedback may ease or frustrate course design and revision decisions. Hence, actionability is the extent to which data, an object, or aspect of practice facilitates the ease of flow in situated activity as it happens. Consequently, for those using student feedback to make decisions about a course, actionability of student feedback is a property of situated action within course design practice. Under this view, instructors use their intuition along with student feedback as tools in present activity to generate new approaches to teaching or learning. Because course design decisions are often made some time after activities or by stakeholders not directly involved in the learning and instruction, the ability of student comments to dynamically afford course design decisions by situating stakeholders in activity context through summarized and individual comments is critical.

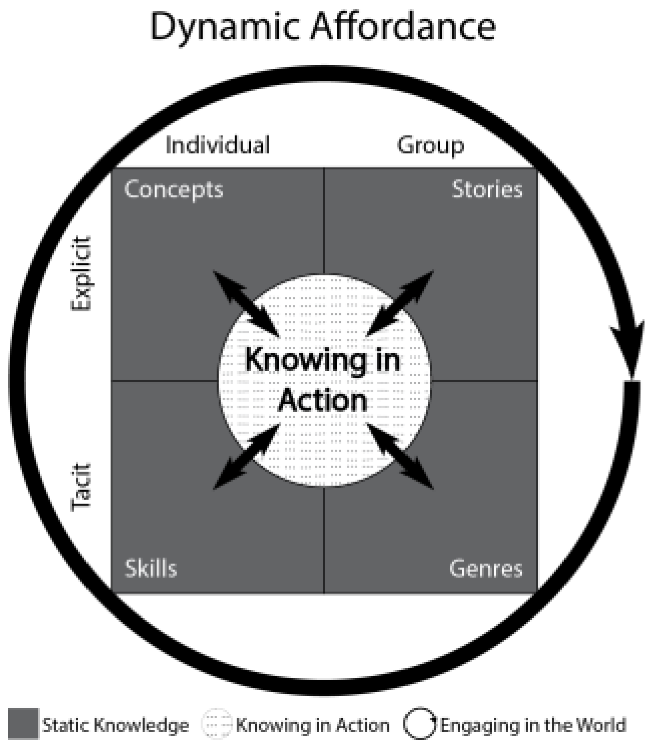


Figure 1. Generation of meaning through dynamic affordance, adapted from Cook and Brown (1999)

Although we did not test the actionability construct (See Figure 1) in course revision practices directly in this study, we used the definition to test whether the affordances of real-time feedback were plausibly actionable. We identified actionable feedback by the criteria in Table 1. For example, the title of a feedback, could be *Module 5 Feedback*, as in the current study. A real-time feedback's meta data would indicate that its submission was submitted close to when the learning activity deliverable was submitted. No assignment submission times were collected, but because the subsequent feedback submission times from individual students were not clumped together, we assumed each was submitted immediately after module completion. Criteria 1-3 would be met by such indicators. For example, one commenter from the study wrote "Also, the "breadcrumbs" did not navigate me back to Module One when completing my profile or when posting to the "Something New" discussion. Perhaps, I was doing something wrong or looking in the wrong place. But, I am new to this platform and that is to be expected." The sentiment was negative, because the course site navigation is unclear. Criteria 4-6 were met by the comment and it would be categorized as actionable because it indicated what exactly about a specific page was confusing. Where as the comment, "So far so good." Would not be actionable as it did not meet criteria 5-6.

Table 1

Actionability Criteria

Criterion	Criterion Description
1.	Demonstrates that it was collected during or near the time of the activity
2.	Names or indicates which learning activity
3.	Indicates when the learning activity happened (by sequence or time)
4.	Reveals the student's sentiment regarding the activity
5.	Identifies the topic(s) of the issue(s) at hand
6.	Indicates why the issue(s) eased or frustrated the activity

In addition to individual comment actionability, we also propose that when the number of comments becomes sufficiently representative of a topic, real-time feedback would be plausibly actionable even if its individual comments may not all be actionable. Real-time feedback comments are likely to be actionable because they afford the situated meaning of the students' experience, concretely captured in the learning activity. In turn, stakeholders making course design decisions then have sufficient context to make meaningful course revisions.

Research Questions

In order to begin the examination of our theory we analyzed real-time feedback that learners gave throughout 16 sections of an online course, as well as compared students' real-time feedback with their end-of-course feedback. The research questions were:

1. To what extent did the course participants provide real-time feedback throughout their course?
2. What was the nature of real-time feedback?

3. What factors of the feedback are related to comment actionability?
4. Does real-time feedback result in more actionable student feedback than end-of-course feedback?

Method

This study examined real-time and end-of-course feedback that students gave during an online, 2-week professional development course. Even though the online course was only 2 weeks, we felt real-time feedback at the activity-level of the course would be relevant no matter the length of the course. However, we postulated that comparisons between real-time feedback and end-of-course feedback could be affected by a potential absence of recall effects (Estelami, 2014; Nicol & Macfarlane-Dick, 2006) due to the relatively short time between the activities in question and the end of the course. The course had 9 content modules and 15 sections of students. Each section had 2 facilitators to manage the course. Course sections started on Mondays throughout October and November of 2014. The students were composed of 276 instructors, staff, and administrators from 99 education institutions. No age or other demographic data was collected.

Materials

Each student completed the 9 modules of the course, which included readings, discussion forums, videos, project-based work, and quizzes. DropThought, a commercially available student feedback platform created by the first author, was used to collect feedback on each module throughout the course (See Figure 2). DropThought feedback forms consisted of a single, open-ended text feedback form prompt, (e.g., “Module 1 Feedback”) followed by a 4-point, “overall experience” rating scale on the same page, made of labeled smiley face buttons:

Excellent, Good, Fair, Poor (see Figure 2). A survey utility in Moodle, the learning management system (LMS) for the course, was used to collect end-of-course surveys from the students. The survey had 11 rating questions, two qualification questions, and one additional comments question.

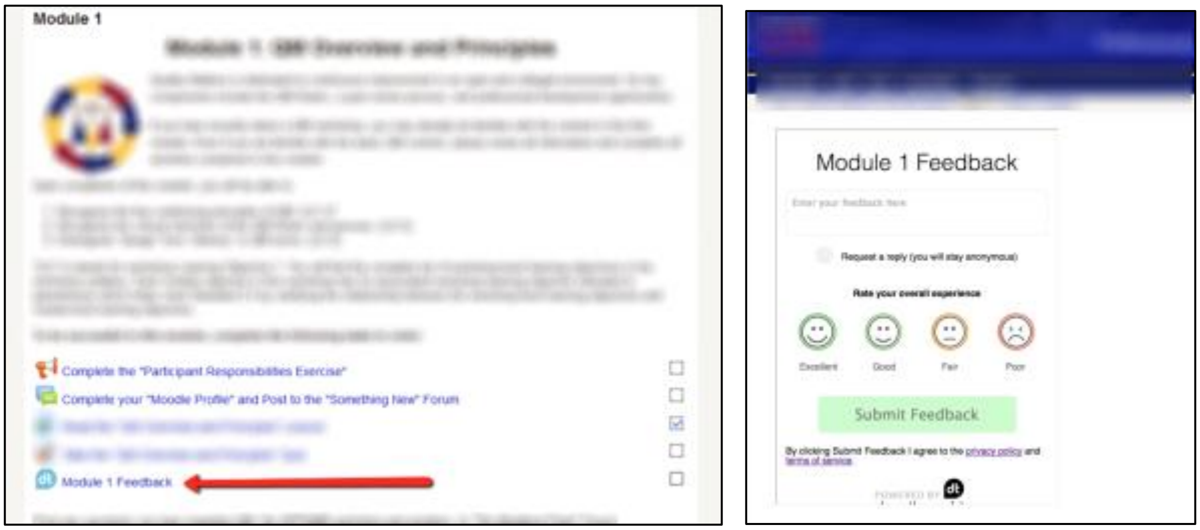


Figure 2. DropThought launch link and feedback form in the Moodle-hosted course

Data Collection Procedure

For real-time feedback, links to launch DropThought feedback forms were added to the end of each course module before the course started (See Figure 2). Facilitators were instructed by their organization to make a course announcement to their course section participants introducing DropThought. The announcement explained that the organization wanted anonymous feedback on individual course modules from course participants, but that the feedback was optional. Facilitators received DropThought dashboards for reviewing and responding to the participants' DropThought feedback messages, and the organization was provided a master dashboard to manage the feedback across all sections. Lastly, electronic end-of-course survey invitations were sent by email to students upon their completion of the 2-week course.

Data Preparation Procedure

In order to analyze quantitative relationships among the qualitative data we first had to create hierarchical categories of topics found in the comments and tally them.

Topic category development. For category development, we analyzed other category lists from literature (Alhija & Fresko, 2009; Braskamp, Ory, & Pieper, 1981; Brockx, Van Roy, & Mortelmans, 2012) and the Quality Matters Rubric Standard (2014), due to its relevance to course quality assurance. Table 2 shows the categories borrowed from these various frameworks.

Table 2

Atkisson (2015) Student Comment Categories

Topic	Subtopic	Braskamp (1981)	Alhija (2007)	Brockx (2012)	QM (2014)	Atkisson
Course or Module	General/Overall	X	X			X
	General		X			X
	Quantity					X
	Student Interest		X			X
	Contribution to learning		X	X		X
	Relevance					X
	Difficulty					X
	Other					X
Evaluation	General/Overall	X				X
	Quantity					X
	Student Interest					X
	Contribution to learning					X
	Relevance			X		X
	Difficulty			X		X
	Method					X
	Content validity					X
	Fairness			X		X
	Other					X
Assignments /Activity/ Content	General/Overall					X
	Quantity	X	X	X		X
	Student Interest		X			X
	Contribution to learning		X	X		X
	Relevance		X			X
	Difficulty		X	X		X
	Other			X		X
Syllabus	Learning Objectives	X	X	X		X
	General					X
	Other					X
Student Composition	Class size		X			X
	Heterogeneity		X			X
	Discipline problems		X			X
	Overall academic level		X			X
	Overall motivation		X			X
Instructor	General/Overall	X	X	X		X
	Subject expertise	X	X			X
	Rapport, Person	X	X	X		X
	Flexibility and consideration	X	X			X
	Interest/Dynamism /Enthusiasm/ Commitment	X	X	X		X
	Other	X				X
Teaching	General/Overall					X
	Improvement	X				X
	Instructor-student	X	X			X
	Clarity		X			X
	Organization	X		X		X
	Environment (created)					X
	Contribution to learning		X			X
	Activating students		X	X		X
	Supervision quality		X			X
	Use of course content		X			X
	Other					X
Logistics	Navigation				X	X
	Usability				X	X
	Instructions				X	X
	Accessibility				X	X
	Technology				X	X
	External resources				X	X
Other	General					X

Content classification into categories. We classified student comments into categories by topic, subtopic, sentiment, actionability, and purpose (see Table 3). Hierarchical topic classification was needed because across student comments, common phrases or words that describe qualities of objects or experiences often appear, but about different topics. For example, some students may describe unclear activity instructions with the word, "clarity," while others may use the same word when mentioning poorly explained concepts. Hence both the "instructions" and "concepts" topics had a subtopic of "clarity." Stakeholders in the course design need to know, not just that clarity was mentioned, but also about which objects or activities it was associated with so they can make specific design changes in the course. Sentiment in this study refers to a 3-point scale, positive, neutral, negative. We gave each comment topic a sentiment category to identify which topics and subtopics were points of frustration or effectiveness for the students. Actionability refers to whether the comment provides enough context to make a course design decision as a result. Only specific comments as defined previously serve as an index for the comment reader to understand the student's situated experience. Comment purpose refers to the argument structure of the comment, whether it was a statement, question, suggestion, or request. We anticipated that questions, suggestions, and requests would be more likely to be actionable than statements.

Table 3

Real-Time and End-of-Course (EOC) Feedback (FB) Variables and Descriptions

Column Name	Variable Description	Real-Time	EOC
FB ID	Unique identifier for each FB	Yes	Yes
FB Sub ID	Incremental number for each comment's topic break	Yes	Yes
Person*	Participant's email ID	Yes	Yes
Course Section	One of 15 course section numbers	Yes	Yes
Course Module	One of 9 course modules	Yes	NA
FB Date	FB date submitted	Yes	Yes
FB Time	FB time submitted	Yes	Yes
FB Comment	Full participant comment on course module or EOC	Yes	Yes
FB Part	Comment parsed into topics, one row per topic	Yes	Yes
Topic	Assigned topic category to comment part	Yes	Yes
Subtopic	Assigned subtopic category to comment part	Yes	Yes
Sentiment	Assigned sentiment rank (1,0,-1) to comment part	Yes	Yes
Actionability	Assigned level of specificity (General, Specific) to comment part	Yes	Yes
Purpose	Assigned purpose of comment (Statement, Question, Request, Suggestion)	Yes	Yes

* Unique identifiers were anonymized

Classification procedure. Next, the first rater parsed 2,092 comments into sub phrases and sentences based (parsings) on when the topic of the comment changed. For example, in the feedback, "All information and activities were easy to find. Content was relevant and useful. A great start," each sentence had a different topic and was parsed per sentence. Then the first rater assigned each comment segment to a topic and subtopic, or the "other" topic if it did not fit any other categories. Levels of sentiment and types of requests were also coded for each parsing (see Table 3). A second rater tagged 10 percent of the topic phrases within each module. Phrases were selected at random by a number generator. Interrater reliability was tracked and Kappa values were calculated for the categories and sentiments (Table 4). Kappa agreement values ranged from Moderate to Substantial (Approximate p=0.000), except for Real-Time Subtopic (Kappa=0.2430, Approximate p=0.000) and End-of-Course Subtopic (Kappa=0.2680, Approximate p=0.000), (Landis & Koch, 1977). A total of 2,092 comments were parsed into

2,970 comment parts (parsings) and classified into 8 topics, 27 subtopics, four communication type categories, three sentiment-level ordered categories, and one dichotomous dependent variable.

Table 4

Kappa Measure of Interrater Reliability Agreement

Topic	Feedback Type	Kappa	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Actionability	Real-Time	0.5540	0.0620	10.401	0.000
Purpose	Real-Time	0.4800	0.1280	8.462	0.000
Topic	Real-Time	0.6670	0.0360	17.596	0.000
Subtopic	Real-Time	0.2430	0.0220	19.708	0.000
Sentiment	Real-Time	0.6780	0.4200	13.6000	0.000
Actionability	End-of-Course	0.4390	0.1840	2.5430	0.000
Purpose	End-of-Course	0.6460	0.3240	3.3130	0.000
Topic	End-of-Course	0.6690	0.1450	4.5980	0.000
Subtopic	End-of-Course	0.2680	0.2680	6.2180	0.000
Sentiment	End-of-Course	0.7550	0.1200	4.0830	0.000

a. *Not assuming the null hypothesis.*
b. *Using the asymptotic standard error assuming the null hypothesis*
Agreement: Fair, 0.21–0.40; Moderate 0.41–0.60; Substantial 0.61–0.80 (Landis & Koch, 1977)

Data

DropThought feedback messages were matched per student to exported participant end-of-course survey responses by the participant's email and/or name. Once the data sources were linked, participant and institution identifiers were anonymized with unique numbers. Table 3 shows the total set of variables used in the analysis.

Question 1 data. To what extent did the course participants provide real-time feedback throughout their course? We examined how much real-time feedback students gave by frequency counts, and percentages.

Question 2 data. What was the nature of real-time feedback? Comment parsings that met the Actionability criteria in Table 1 were counted as actionable. We observed actionability of student's feedback comments by count and percentage of comment parsings. We also grouped observations by topic and sub topic, sentiment, and statement purpose to see which types of comment parsings were most usually or correlated with actionability.

Question 3 data. What factors of the feedback are related to comment actionability? In addition to descriptive statistics and correlations regarding the nature of actionable feedback, we also wanted to know what factors of student comments were predictive of whether feedback parsings were actionable. Given the effect of tweet authorship on tweet sentiment in online learning (Atkisson, n.d.), we wanted to know if the actionability variance was explained more by the grouping variable, comment authorship (see Equation 4), or by the grouping variable plus comment property predictors such as subject, sentiment, etc. across individuals (see Equation 7). The comparison of these two hierarchical models identifies which factors contribute most to whether a student's comment is actionable.

First, we formulated a hierarchal generalized linear (HGLM) random intercept model (see Equation 2) for a dichotomous outcome variable with no predictors to see whether comment parsings written by the same author tend to be more actionable than those of other authors. This gave us a baseline for the dependent variable of actionability that we used to compare with the second model that had predictors (various comment properties). HGLM was used to account for the dichotomous probability distribution in the dependent variable, which is log based, not a normal distribution (Heck, Thomas, & Tabata, 2013). Usually hierarchical modeling in education is used to account for students grouped into classes. In this case, however, we used it to examine to what extent the variance in actionability of a comment (observed at level-one) was attributed to authorship variance at level-two.

Accounting for groupings in the observed variables was important because ordinary least squares regression (OLS) assumes each observation is independent. Groupings in the data violate that assumption and would result in overestimated statistical significance if multiple levels were not used in the analysis (Heck & Thomas, 2008). Intra-class correlation (ICC) can be used to evaluate whether within groups variance has a significant effect (Heck & Thomas, 2008). In cases where cluster sizes vary significantly, however, Muthén and Satora's (1995) design effect (DEFF) should be used, where c is the average cluster size and p is the ICC (Equation 1). A DEFF value of 2 or greater indicates that a multilevel regression analysis should be used because of a significant clustering effect, (e.g., comment authorship) in the data (Satorra & Muthen, 1995).

$$DEFF = 1 + (c-1)p \tag{1}$$

Equation 2 is an HGLM with a dichotomous outcome variable, actionability. The predicted quantity was the log of the odds of a student's comment parsing i being actionable for the jth student. Let Y_{ij} = 1 if the ith comment part classification is actionable for student j and Y_{ij} = 0 otherwise. Let Y_{ij} be the probability that Y_{ij} =1, which varies randomly across students. There were 2,970 comment parsings.

$$\text{logit}(Y_{ij}) = \beta_{0j} \tag{2}$$

The level one equation has no error term because it was accounted for in the link function (O'Connel et al., 2008). The link function is used to map the result from a dichotomous distribution to an estimated normal distribution (Heck et al., 2013). One comment parsing was a reference, while the weights of the remaining comment parsings were relative to it, hence q = 1,…, x-1, for a total of 2,970 comment parsings. We specified a random intercept model in anticipation that variability in intercept means would vary at both levels of the model. This implied at level-1 that the likelihood of a comment's actionability would vary across a student's comments. At level-2, the mean of a student's comment actionability would vary across students. In the context of our study, the level-2 terms are:

= dependent variable, log odds measure of Actionability for the *i*th sub-comment at level-1 for *j*th student at level-2

β_{0j} = Actionability intercept for comment part i for student j that is not related to the model's residual error

The level-2 equations are:

$$\beta_{0j} = \gamma_{00} + \zeta_{0j} \tag{3}$$

where $q = 1,\ldots, x\text{-}1$. For this study, slopes are not random and are assumed to be equivalent across students. In the context of our study, the level-2 terms are

β_{0j} = intercept for the *j*th level-2 unit (student);
γ_{00} = overall mean intercept;
ζ_{0j} = random effects on the jth level-2 unit on the intercept.

All the random effects are considered to be normally distributed with means of zero and unknown variances (Kamata, Bauer, & Miyazaki, 2008). The combined equation is the following:

$$\text{logit}(Y_{ij}) = \gamma_{00} + \zeta_{0j} \tag{4}$$

The second univariate, HGLM tested to what extent actionability was due to various comment properties, including topic, sub-topic, sentiment, and comment purpose, in addition to comment authorship. In other words, we wanted to see which topics or comment types were more or less actionable given the effects of student authorship.

In the second model for question three, we anticipated that the probability of a comment's actionability would be related to the module in which it was written and to its other comment attributes such as topic, sub topic, sentiment, and purpose.

Equation 5 is a HGLM with a dichotomous outcome variable, actionability.

$$\text{logit}(Y_{ij}) = \beta_{0j} + \sum_{1}^{x-1}\beta_{1j}A_{1ij} + \sum_{2}^{x-1}\beta_{2j}G_{2ij} + \sum_{3}^{x-1}\beta_{3j}Z_{3ij} + \sum_{4}^{x-1}\beta_{4j}W_{4ij} + \sum_{5}^{x-1}\beta_{5j}M_{5ij} \quad (5)$$

Again, the level-one equation had no error term because it was accounted for in the link function (O'Connel et al., 2008). One comment parsing was a reference, while the difficulties of the remaining comment parsings were relative to it, hence $q = 1,\ldots, x\text{-}1$, for a total of 2,951 comment parsings. Topic (8 nominal categories), Subtopic (29 nominal categories), Sentiment (3 ordered categories), and Purpose (4 nominal categories) were comment part attributes specified as predictors of the comment's actionability at level-one and two. The Module (9 nominal categories) in which the comment was written was also specified as a predictor at level-one and two. The level-two terms are listed in Table 5.

Table 5

Null HGLM Level-1 (Observed Parsings) Coefficients and Descriptions

Coefficient	Description
$\text{logit}(Y_{ij})$	Dependent variable, log odds measure of actionability for the *i*th parsing at level-1 for *j*th student at level-2;
A_{1ij}	parsing Topic for sub-comment *i* for student *j*;
G_{2ij}	parsing Subtopic for sub-comment *i* for student *j*;
Z_{3ij}	parsing Sentiment for sub-comment *i* for student *j*;
W_{4ij}	parsing Purpose for sub-comment *i* for student *j*;
M_{5ij}	parsing Course Module for sub-comment *i* for student *j*;
β_{0j}	Actionability intercept for parsing *i* for student *j* that is not related to the model's five predictors;
β_{1j}	regression coefficient associated with Topic relative to level-2 intercept, student *j*;
β_{2j}	regression coefficient associated with Subtopic relative to level-2 intercept, student *j*;
β_{3j}	regression coefficient associated with Sentiment relative to level-2 intercept, student *j*;
β_{4j}	regression coefficient associated with Purpose relative to level-2 intercept, student *j*;
β_{5j}	regression coefficient associated with Course Module relative to level-2 intercept, student *j*;

The level-two equations are:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}A_j + \gamma_{02}G_j + \gamma_{03}Z_j + \gamma_{04}W_j + \gamma_{05}M_j + \zeta_{0j}, \\
\beta_{1j} &= \gamma_{10} + \zeta_{1j}, \\
\beta_{2j} &= \gamma_{20} + \zeta_{2j}, \\
\beta_{3j} &= \gamma_{30} + \zeta_{3j}, \\
\beta_{4j} &= \gamma_{40} + \zeta_{4j}, \\
\beta_{5j} &= \gamma_{50} + \zeta_{5j}
\end{aligned} \tag{6}$$

where $q = 1,\ldots, x\text{-}1$. For this study, slopes were random and are assumed to vary between students. The level-two terms are listed in Table 6.

Table 6

Null HGLM Level-2 (Grouped by Student) Coefficients and Descriptions

Coefficient	Description
β_{0j}	intercept for the *j*th level-2 unit (student)
$\beta_{1\ldots5j}$	slope for the *j*th level-2 unit (student)
A_j	parsing Topic for student *j*;
G_j	parsing Subtopic for student *j*;
Z_j	parsing Sentiment for student *j*;
W_j	parsing Purpose for student *j*;
M_j	parsing Course Module for student *j*;
γ_{00}	overall mean intercept adjusted for model's five predictors;
$\gamma_{10\ldots50}$	overall mean intercept adjusted for model's five predictors;
γ_{01}	regression coefficient associated with Topic relative to level-2 intercept (student);
γ_{02}	regression coefficient associated with SubTopic relative to level-2 intercept (student);
γ_{03}	regression coefficient associated with Sentiment relative to level-2 intercept (student);
γ_{04}	regression coefficient associated with Purpose relative to level-2 intercept (student);
γ_{05}	regression coefficient associated with Course Module relative to level-2 intercept (student);
ζ_{0j}	random effects on the jth level-2 unit adjusted for the model's five predictors on the intercept;
$\zeta_{1\ldots5j}$	random effects on the jth level-2 unit adjusted for the model's five predictors on the slope.

All the random effects are considered to be normally distributed with means of zero and unknown variances (Kamata et al., 2008). The combined equation is the following:

$$\begin{aligned}
\text{logit}(Y_{ij}) &= \gamma_{00} + \gamma_{10}A_j + \gamma_{20}G_J + \gamma_{30}Z_j + \gamma_{40}W_j + \gamma_{50}M_j \\
&+ \sum_{1}^{x-1}\beta_{1j}A_{1ij} + \sum_{2}^{x-1}\beta_{2j}G_{2ij} + \sum_{3}^{x-1}\beta_{3j}Z_{3ij} + \sum_{4}^{x-1}\beta_{4j}W_{4ij} + \sum_{5}^{x-1}\beta_{5j}M_{5ij} \\
&+ \zeta_{1j}A_j + \zeta_{2j}G_j + \zeta_{3j}Z_j + \zeta_{4j}W_j + \zeta_{5j}M_j + \zeta_{0j}
\end{aligned} \tag{7}$$

Lastly for question three, we attempted to compare which model had a stronger effect on frequency of actionable comments but were limited by available analyses for HGLMs.

Question 4 data. Does real-time feedback result in more actionable student feedback than end-of-course feedback? To test whether actionable comments were more frequent real-time feedback than in end-of-course feedback comments, we compared the proportion of comments that were actionable. Because the end-of-course comments were further away in time from a student's course activity and focused on the course overall rather than at the course's activity level, we anticipated the end-of-course feedback would have a lower percentage of actionable comments than real-time feedback. For only the students who gave comments in their end-of-course feedback, we tested this hypothesis with a *Z*-test of the proportion of actionable comments in real-time feedback with the actionable proportion of end-of-course comments.

Results

Several researchers have called for more formative feedback research (Beaty, 1997; Dennen & Bonk, 2007; Hendry et al., 2001; Jahangiri, Mucciolo, Choi, & Spielman, 2008; Jara & Mellar, 2010; Lewis, 2001; Wagner et al., 2015; Woloschuk et al., 2011). Yet, little research has been done on real-time feedback. Atkisson and West (n.d.) found that only 2 out of 100 reviewed studies had collected feedback more often than weekly. This study exhibits a novel, per module collection of feedback in an accelerated, 2-week course. Thus, questions about the nature of real-time feedback from student to instructor are important to answer.

Question 1 Results

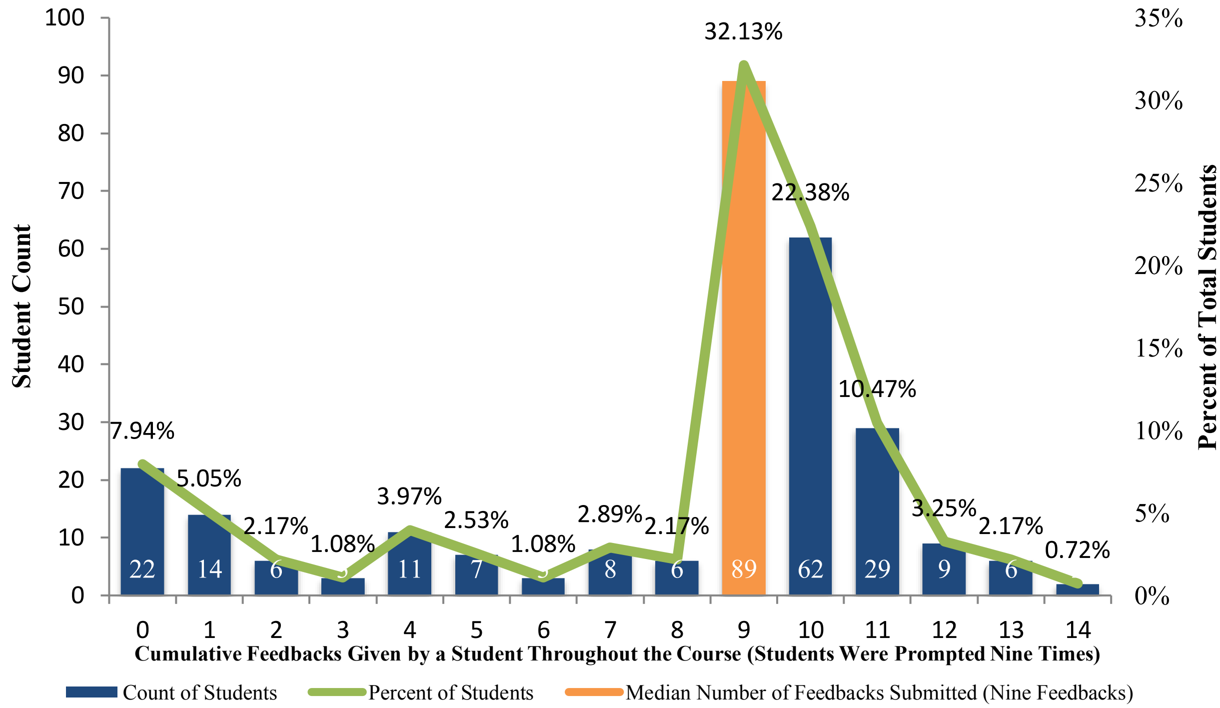
To what extent did the course participants provide real-time feedback throughout their course? We found that even though the feedback was optional, the course was accelerated, and the audience was made of working professionals, the clear majority of students left a DropThought style feedback that at least had a smile rating (92%, See Table 7). Across the 15 sections of courses, 87% of students left at least one real-time feedback with a comment when prompted at the end of each course's content modules. Furthermore, Figure 3 shows that over 70% of the students gave 9 or more feedbacks, with a median of 9 and a range of 9 to 14. Only 13% of students gave zero or one feedback per course. These participation rates were much higher than comparative end-of-course survey rates from the literature. For example, Nulty (2008) found across 9 studies that online end-of-course response rates ranged from 23% to 47% with an average of 33%. At a minimum, real-time feedback appears to afford a significant advantage over end-of-course feedback in the practice of making course revisions by engaging a much broader sample of students and deeper set of student touch points throughout a course.

Table 7

Student Counts and Percentages by Feedback Type Across 15 Course Sections

Students by Feedback Type	Student Count	Percent of Total Students
Real-Time Feedback	253	91.67%
Real-Time Feedback with Comment*	240	86.96%
Total Students	276	100.00%

*Note: *DropThought requires the written comment. Thirteen students wrote "no feedback" to bypass the required comment and submitted a rating.*



Note. Figure 3 shows the total count of feedbacks that a student gave throughout the 2-week course. For example, the zero category on the x-axis represents 22 students or the 7.94% of total students that did not give any real-time feedback throughout the course. Feedback was requested 9 times, but not required. It is of note that the median feedbacks submitted was 9, and over 70% of students gave 9 feedbacks or more.

Figure 3. Count and percent of students by number of real-time feedbacks given

Question 2 Results

What was the nature of real-time feedback? Even though an understanding of real-time feedback response rates of the observed students is an important addition to the formative feedback cannon, it did not move beyond the granularity claims from the formative feedback literature discussed earlier. Hence, we proposed text analysis by categories as an affordance of the real-time feedback data for course quality stakeholders, because patterns are not identified or confirmed by looking at individual comments alone.

Overall, 253 of 276 students gave 2,092 real-time feedbacks throughout the course. Each feedback consisted of a rating and an open-ended comment. We broke up the comments into parsings, phrases by topic, which resulted in 2,970 parsings, or comment parts. Seventeen percent or 511 of all parsings were actionable. Parsings were classified into 8 topics and 27 subtopics, which ranged from 17 to 1 subtopics per topic (See Table 8).

Because over 70% of the students commented on every content model at least once, there was a wide variety of comments. Overall, however, tactical-related themes rose to the top of actionability within the top-level topics. Whereas, learning-, interest-, and relevance-related themes were largely at the lower end of actionability within top-level topics. Also in Table 8, the topics for the activity-level comments on the course, (e.g., Evaluation, and Logistics) had higher percentages of actionability, 37% to 63%. Whereas, topics for higher-level themes or organization in the course, (e.g., Module and Course) had lower percentages of actionability, 5% to 7%.

For example, Table 9 shows a frequent comment submission, “I found this module to be helpful and well organized” that painted how students felt about particular modules (known from the metadata), but it is not particularly actionable. On the other hand, “Instructor’s e-mail should link from their introductions. Announcement link needs to be clear – the start of modules” was one of many Logistics topic comments that provided sufficient context to be actionable as an individual comment.

Even so, for course redesign stakeholders, individual comments are not the only way to find actionable results. For example, the subtopics of Time and/or Quantity, which referred to the large amount of work for the short amount of course time, (e.g., “Three weeks would be a much-improved time frame in my view, instead of 14 days.”), were top-4 subtopics for actionability across the 5 topics with the most comments (Module, AAC, Evaluation, Course, and Logistics). Many students noted the underestimated time required for the course at the assignment, module, and course levels. Nevertheless, even if individual comments about time or workload were not actionable (as most were like the example in Table 9), such as "This module took a while to complete", the large pattern (N=193) of time mentions across students would be an actionable data point in aggregate for course quality stakeholders. Hence, we found that specific comments captured in real-time were plausibly actionable, not only at the individual feedback level, but also through simple aggregations of like topics and themes once they were classified.

Table 8

Subtopic Descriptive Statistics by the Five Highest-Commented Topics, Sorted Descending by Actionability

Category	Module			
	Subcat.	Parsings	Actionable	Sentiment
	Time	87	22	-0.8161
	Quantity	75	12	-0.2133
	Design	51	6	0.7059
	Contribution to Learning	313	3	0.9233
	Instructions	22	3	-0.5455
	Difficulty	58	2	0.0690
	Usability	8	1	0.5000
	General	238	0	0.8655
	Student Interest	104	0	0.8942
	Learning Objectives	5	0	0.6000
	Other	8	0	0.5000
	No Feedback	32	0	0.0000
	Quality	55	0	0.9636
	Relevance	20	0	0.9000
	Navigation	7	0	0.4286
	Content Validity	2	0	0
	Accessibility	1	0	1
	External Resources	1	0	1
Total	18	1,087	49	0.5667

Category	Activity, Assignments, Content			
	Subcat.	Parsings	Actionable	Sentiment
	Design	107	56	-0.1869
	Time	84	34	-0.7976
	Instructions	66	29	-0.6364
	Content Validity	47	24	-0.4255
	Quantity	32	12	-0.6563
	Relevance	51	9	0.0000
	Accessibility	10	9	-0.8000
	Usability	12	7	-0.8333
	Difficulty	30	6	-0.5000
	Navigation	5	4	-0.6000
	Student Interest	235	3	0.9362
	Quality	105	3	0.9048
	Contribution to Learning	313	2	0.9169
	Learning Objectives	1	1	-1
	External Resources	3	1	-0.3333
	Other	5	0	0.0000
	General	11	0	0.9091
Total	17	1,117	200	0.361683

Category	Evaluation			
	Subcat.	Parsings	Actionable	Sentiment
	Content Validity	52	38	-0.8846
	Relevance	17	6	-0.2353
	Design	18	5	0.3333
	Quantity	7	5	-0.7143
	Usability	6	4	-0.5000
	Instructions	8	4	-0.6250
	Time	5	3	-0.6000
	Navigation	4	3	-0.5000
	Difficulty	14	2	0.1429
	Quality	11	2	0.6364
	Contribution to Learning	36	1	0.6389
	Other	4	1	0.2500
	Student Interest	7	0	0.7143
	Fairness	3	0	1
	General	6	0	0.6667
Total	15	198	74	-0.085859

Category	Course			
	Subcat.	Parsings	Actionable	Sentiment
	Time	8	5	-0.7500
	Design	7	1	0.1429
	General	21	0	1
	Quality	6	0	1
	Student Interest	9	0	1
	Learning Objectives	1	0	0
	Contribution to Learning	16	0	0.8750
	Instructions	1	0	1
	Accessibility	1	0	1
	Relevance	6	0	0
	Content Validity	1	0	1
	Quantity	2	0	0
	Other	1	0	1
	Difficulty	3	0	0.3333
Total	14	83	6	0.6125

Category	Logistics			
	Subcat.	Parsings	Actionable	Sentiment
	Usability	76	66	-0.8421
	Navigation	90	47	-0.6000
	Technology	62	28	-0.4839
	Time	9	8	-1
	External Resources	6	5	-0.5000
	Instructions	9	3	0
	Support	1	1	1
	Accessibility	1	1	-1
	Content Validity	1	1	-1
	Design	1	0	-1
Total	10	256	160	-0.63137

Note: Three top-level categories are not shown (Instructor, Students, and Other), because they had a total count of 15 feedback parsing together.

Table 9

Actionable and Not Actionable Category Examples Sorted by Percent Actionable

Category	Topic	Parsing Count	Actionable Count	% Actionable Parsings	e.g. Actionable	e.g. Not Actionable
Topic	Logistics	255	160	63%	instructors e-mail should link from their introductionsAnnouncement link needs to be clear - at start of modules.	I'm brand new to Moodle, so I am adjusting and learning.
	Evaluation	198	74	37%	This has a box checking feel to me. Particularly on the last quiz I found the questions to be busywork and not helpful to the learning process.	I thought the exam was well written.
	AAC	1,117	200	18%	The lesson print is very small.	and the underlying principles information was clear and concise.
	Course	80	6	8%	it would be great to prolong the workshop to 3 weeks.	Highly rewarding overall experience
	Module	1,087	49	5%	This was a rather long module... Took more time than the other modules did, and I wasn't prepared for that.	I found this module to be helpful and well organized.
Subtopic (11 Most Actionable Subtopics)	Accessibility	13	10	77%	One feedback: the 'playing cards' for the Design and Delivery differences, include 'visual blinkers' with the audio sounds.	Just as a test, I also viewed the closed captioning for the module. Some of the text is interesting.
	Usability	102	78	76%	This is my second attempt at feedback.	Some of the multiple choice pull down questions took a bit to figure out. Confusing to figure out at first.
	Content Validity	103	63	61%	I think Question #10 is misleading based on how it is worded in the section "Important Reminders About the Course Worksheet."	I found this standard to be subjective and
	External Resources	10	6	60%	I think supplying a website with "searchable" Bloom's words would be helpful.	I also appreciate the tip on the captioning software, [instructor name] I may have to try that...
	Navigation	106	54	51%	It was a bit confusing to find the link to the "something new" forum (despite the instructions provided).	I don't like the navigation of the Moodle work/posts.
	Responsiveness	12	6	50%	I asked for assistance in the Muddiest Points forum and got a partial response. I posted again and still have NO RESPONSE.	I appreciate the timeliness of our two instructors, especially over the Thanksgiving break!
	Technology	62	28	45%	I've just watched the video and found the audio quality to be poor.	had some difficult with moodle's forum.
	Time	193	72	37%	However, I do not like the waiting time for my post to be shown and I need to wait to see others' posts.	This Module took awhile to complete.
	Design	183	68	37%	I would like to see the course after recommended improvements have been made. [Note: Course Evaluation Activity]	Nice structure
	Instructions	106	39	37%	To provide instructions, step-by-step on posting to the discussion board forum for the first time.	Challenging to know how much detail is expected.
	Quantity	116	29	25%	I might have split this into two modules. Perhaps learning the rubric itself should be a stand-alone module.	It was long and tedious.
Sentiment	-1	846	437	52%	and I don't like that I can't click "next" to read the next post; rather, I have to click the "back" arrow.	The introduction to moodle was difficult to learn.
	0	179	47	26%	More examples and application of Standard	The amount of content provided was enough.
	1	1,742	12	1%	This is great information and needs to be one of the first modules presented. I now understand why this course is training me the way it is now.	I have learned a lot from the workshop that can be applied to my courses.
Purpose	Request	4	4	100%	Please reply to my previous feedback.	[No Example]
	Suggestion	172	162	94%	Some cue like " review these principles" and then take the quiz would be helpful. If it's there, I didn't see it.	It would be helpful to include this type to assignment in other areas of the course to cover all learning styles...Just a thought!
	Question	22	16	73%	When will we know our scores on the Module 2 Quizzes?	The time investment is growing? Do course reviews get quicker as you become more familiar with the process?
	Statement	2,569	314	12%	Thanks for not including a lot of busy-work (one reply to a peer instead of two, for example).	This was an excellent course that will help me with my on-line courses.

Tables 8 and 9 illustrate what types of comments were better for actionable design decisions. As we analyzed the text categories with correlations, however, we could rule out a variety of potential effects on Actionability, which included the number of students who submitted feedbacks, the number of feedbacks submitted, the number of resulting parsings overall and when grouped by course module, feedback purpose (i.e., request, suggestion, question, statement), or the parsings' topics and subtopics. On the other hand, we found a strong inverse relationship between actionability (the average percent of parsings) and sentiment (average parsing sentiment rating on an ordinal integer scale from 1 to -1) among six groupings of comment parsings (See Table 10), Purpose, Module (course content section), Topic, Subtopic Evaluation, Subtopic Activity-Assignment-Content (AAC), and Subtopic Module (mentioned in comment).

Table 10

Correlation Between Actionability and Sentiment per Category

Category	*r* Type	*r*	r^2	df	Sig.
Purpose	Spearman	-0.9764	0.9533	2	0.0236
Module (Course Content)	Spearman	-0.9359	0.8760	7	0.0002
Topic	Spearman	-0.6895	0.4754	6	0.0399
Evaluation	Spearman	-0.8857	0.7844	13	0.0000
AAC	Spearman	-0.8212	0.6744	15	0.0001
Module (Comment Topic)	Spearman	-0.5649	0.3191	17	0.0117

The strongest Actionability-Sentiment correlation related to the purpose of the students' comment parsing (See Table 10). Purpose consisted of 4 categories, Request (N=4), Suggestion (N=178), Question (N=23,), and Statement (N=2,765). See Table 11 for descriptive attributes and Table 9 for text examples. Purpose had a near perfect inverse relationship between actionability and sentiment (r_s=-0.9764, p=0.236) with more than 95% of the variance explained by the model. The greater proportion of actionable comments the lower likelihood of a negative comment sentiment. Communication types, though not ordinal, ranged from actionable and negative to not actionable and positive in the following order, Request, Suggestion, Question, and Statement. Essentially, if instructors are seeking actionable comments that afford course design decisions, it appears they may want to avoid prompting students for statements alone.

Table 11

Real-Time Feedback Descriptive Attributes by Student's Feedback Communication Purpose

Fb Purpose	Persons	Feedbacks	Parsings	Actionable	% Actionable Parsings	Avg. Sentiment
Request	4	4	4	4	100%	-1
Suggestion	93	162	178	167	94%	-0.67
Question	20	21	23	16	70%	-0.26
Statement	255	2,013	2,765	324	12%	0.42
Grand Total	255	2,092	2,970	511	17%	0.35

Note: Purpose refers to the purpose of the observed comment parsing to the instructor. Sentiment is an ordinal scale from 1 to -1. Each parsing received a sentiment score that has been averaged per category.

Actionability and Sentiment were also inversely correlated across the course modules (r_s= -0.9359, p=0.0002) with 88% of the variance explained by the model (See Table 10). Module referred to the 9 feedback prompts given to students at the end of each course content module (See Table 12). The greater proportion of actionable comments, the lower the likelihood of positive comment sentiment. The greatest percent of actionable parsings was in the middle of the course, culminating in Module 4, at 26% of 328 parsings, coupled with the second most negative average parsing sentiment among modules (0.05 on a scale of 1 to -1). The lowest actionable parsings percentage among modules was Module 9, at 5% of 277 parsings, which also had the highest average sentiment (0.78 on a scale of 1 to -1). Actionability and Sentiment, however, showed a downward trend in sentiment and upward trend in actionability for modules two through six. Crossing feedback topics with the course content modules revealed that the quantity of work for the amount of time and Moodle navigation and usability were related to the bump in actionability for modules two through six. Beginning- and end-course modules had higher proportions of non-actionable comments about the course overall, which as described earlier, are highly related to positive feedback. In terms of real-time feedback for course revision practice, it appears that the beginning and end of the course resulted in more general and positive feedback, while in between we saw higher percentages of actionable feedback. Hence instructors and designers may want to consider more specific feedback prompts in intro and closing course content modules if module specific feedback is desired. Testing whether such modifications to feedback prompts in fact increase likelihood of actionability at the beginning and end of courses would be an important follow up study.

Table 12

Real-Time Feedback Descriptive Attributes by Course Module

Feedback Prompt	Persons	Feedbacks	Parsings	Actionable	% Actionable Parsings	Avg. Parsing Sentiment
Module 1	238	262	441	87	20%	0.38
Module 2	230	241	317	66	21%	0.30
Module 3	226	247	342	68	20%	0.34
Module 4	220	233	328	86	26%	0.05
Module 5	213	222	340	81	24%	-0.07
Module 6	218	228	344	48	14%	0.37
Module 7	217	229	302	21	7%	0.60
Module 8	216	224	279	39	14%	0.50
Module 9	202	206	277	15	5%	0.78
Grand Total	255	2,092	2,970	511	17%	0.35

Note: Module refers to the content module of the course. Students were prompted for real-time feedback 9 times, once at the end of each module. Sentiment is an ordinal scale from 1 to -1. Each parsing received a sentiment score that has been averaged per category.

We also found the inverse correlation between Actionability and Sentiment among observed parsings grouped by their topic (r_s=-0.6895, p=0.0399) with 47% of the variance explained by the model (See Table 10). The correlation was among the eight content topic categories (See Table 13 for descriptive attributes and Table 9 for text examples). Here also the greater proportion of actionable comments, the lower likelihood that a comment was positive (average of parsings labeled 1, 0, or -1). For example, the highest actionable parsings percentage among top-level topics was Logistics, at 63% of 169 parsings, which also had the lowest average sentiment (-0.63 on a scale of 1 to -1). The top subtopics under Logistics were Usability (69 of 79 parsings actionable), Navigation (47 of 90 parsings actionable), and Technology (33 of 69 parsings actionable). The common theme among the comments was users new to Moodle who struggled with navigating among the online course assignments, which included a lot of looking at reference materials to answer questions. Nevertheless, anecdotally, the specifics of the actionable comments within the logistics themes seemed to vary widely. In practical terms for instructors or designers, sub-subtopics may be needed in the future to help course design stakeholders quickly narrow in on specific issues within, for example, Navigation, or Technology.

Table 13

Real-Time Feedback Descriptive Attributes by Feedback Topic

Topic	Persons	Feedbacks	Parsings	Actionable	% Actionable Parsings	Avg. Sentiment
Logistics	132	223	268	169	63%	-0.63
Other	2	2	2	1	50%	-0.50
Evaluation	120	175	198	74	37%	-0.09
Instructor	25	27	29	6	21%	0.45
AAC	221	934	1,151	202	18%	0.38
Course	116	136	163	8	5%	0.69
Module	234	1,006	1,130	51	5%	0.58
Students	10	10	10	0	0%	0.20
Grand Total	253	2,073	2,951	511	17%	0.35

The takeaway here for stakeholders in course revision practice is that no matter the comment topic, actionable comments are overwhelmingly identified as negative (-1 on a scale from 1 to -1) and the opposite is true from comments not labeled as actionable. This is not to say that non-actionable comments (defined as in Table 1) are useless, rather they require more analysis to derive potential actionable insights across comments. Furthermore, these results do not suggest that that positive comments cannot be actionable, but that in similar circumstances they likely would not be actionable as individual comments. Question 2 results also highlighted the potential needs to prompt students for more than simple statements, to call out desired feedback themes in opening and closing course module feedback prompts, and to have a three-level hierarchy for comment analysis to identify groupings of actionable comments within subtopics.

Question 3 Results

What factors of the feedback are related to comment actionability? From the previous research questions, we found through descriptive statistics and correlations that the observed students responded well to real-time feedback and that certain types of real-time feedback appeared to be more frequently actionable than others. In this question, we wanted to know what multiple factors predicted actionability ($p= 0.05$). Given that most students authored 9 or more comments, first, we needed to see whether student authorship influenced actionability likelihood. In other words, were the differences in comment actionability among students due to chance or were feedbacks from the same student more likely or not to be actionable? If the latter case, then understanding which student attributes are associated with actionable feedback and whether those student attributes can be influenced to improve comment actionability likelihood would become an important affordance to examine in future research.

In related research on student comments, Atkisson (n.d.) found that tweet authorship affected tweet sentiment. The tweets were part of a pre-Coursera, Massively Open Online Course (MOOC) that used Twitter for course communication among 6 instructors and many students. Tweet authorship (1,157 students) had a significant effect on sentiment level ($\beta = 0.1265$, $p < 0.01$, $ICC=0.2241$) of the 7,939 tweets on a continuous scale from 5 to -5 (Atkisson, n.d.). Similarly, we wanted to see if feedback authorship in the current study had a similar effect. Therefore, Model 1 (See Equation 4) examined whether some students tended to write more (or less) actionable feedback parsings than others. Furthermore, in Model 2 (See Equation 7), we wanted to know which comment parsing properties from Question 2 predicted the likelihood of actionability when accounting for the student author effect.

Model 1 was a null, dichotomous outcome (Actionability, yes or no) General Linear Model. We specified a two-level model with observed feedback parsings at level-one and grouping of the feedback parsings by the students who wrote them at level-two. For the level-one model fixed effects, the grand mean of actionability across parsings, $\text{logit}(Y_{ij})$, was -1.7 (p=0.0) when holding random effect predictors to a constant of zero. Hence, for every unit increase in $\text{logit}(Y_{ij})$ the predicted odds were multiplied by e^{β} ($2.71828^{-1.7}$), which equaled the odds ratio of 0.11827. In terms of odds ratio, the comment parsings were about 12% more likely to be actionable than not across the population of comment parsings. Converting the log odds to a unit-level probability, an individual comment parsing was 15% more likely to be actionable.

At level-two (See Random Effects in Table 14), the Z-test (Z=5.688, p=0.0) indicated that the intercept of a student's average proportion of actionable comments (β_{0j}) varied from student to student, which supported the need for a hierarchical model. With the wide range in the count of parsings per student (see Figure 3), however, we calculated the Design Effect (DEFF=2.8466) which should be larger than 2 in order to indicate the need for a multilevel model (Satorra & Muthen, 1995). Thus, some students were more likely than others to write comments that were more actionable.

As with binary outcome Generalized Linear Models, residual variance at level-one is set to 1 and cannot be tested (Heck et al., 2013), but an intraclass correlation (ICC = 17.3165%) can still be calculated with an estimated level-one variance of $\pi^2/3$. Therefore, about 17% of the variance of actionability was due to student to student differences rather than chance. This meant that there was sufficient variance to explain through Model 2 with predictors. Again, this outcome is important because it signaled that students did not respond with the same level of actionability from the same feedback prompts due to their individual attributes. Practically speaking, Model 1 results suggests the need to understand how student attributes are related to actionability to test in future studies whether targeting individualized feedback prompts per attribute would raise the likelihood of actionable comments among student types.

Table 14

Model 1: Null, Two-level, Dichotomous Outcome, Generalized Linear Model, Actionability of Student Comment Parsings by Authorship

Parameter	Estimate	Std. Error	t	Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Fixed Effects: Actionable							
Log Odds	-1.7000	0.0760	-22.4170	.	0.0000	-1.8480	-1.5510
Odds Ratio	0.1827	.	.	.	.	0.1574	0.2120
Unit Level Probability	15.4465%	.	.	.	.	0.1360	0.1749
Random Effect Covariances: Actionable							
Log Odds	0.6890	0.1210	.	5.6880	0.0000	0.4880	0.9730
Odds Ratio	1.9917	.	.	.	.	1.6291	2.6459
Unit Level Probability	66.5744%	.	.	.	.	0.6196	0.7257
$\pi^2/3$	3.2899						
Intraclass Correlation (ICC)	17.3165%						
Average Cluster Size (Parsings/Person)	11.6640						
Design Effect (DEFF)	2.8466						

For Model 2, we added predictors to understand which comment parsing properties predicted the likelihood that the feedback parsings would be actionable at both the population average of comment parsings overall, as well as at the per student author average. Topic (8 nominal categories), Subtopic (29 nominal categories), Sentiment (3 ordered categories), and Purpose (4 nominal categories) were specified at level-one to estimate their effect on the actionability likelihood of an individual comment parsing. We also specified them at level-two to estimate their effect on the average actionability likelihood across a student's comment parsings.

In Model 2, several predictors of Actionability were significant. Table 15 shows that predictors with log odds coefficients greater than zero had a positive relationship with the likelihood of actionable comment parsings, and, conversely, those with log odds coefficients less than zero were negatively associated with the likelihood of actionable comment parsings. Note that the odds ratio is the number of times greater or less than the average actionability of comment parsing population, whereas the unit probability is the probability that an individual comment parsing would be actionable. For example, the subcategory Accessibility parsings (N=13) were 9 times more likely to be Actionable than the average comment population (γ_{02} =2.213, p=0.0, e^{β}=9.1431). Whereas, an individual comment about Accessibility had 90% chance of being actionable. Additional studies would be needed to determine if accessibility feedback or other significant topics are always highly actionable or if they were an artifact of this course's subject matter and audience. See Table 9 for text examples.

The subtopics Usability, (N=105), Responsiveness (N=17), and Content Validity (N=104) all had high odds ratios (e^{β}=5.3069, 4.5997, 2.6778), meaning that parsings from those subtopics were 2.6 to 5.3 times more likely to be actionable in a student's comments than the level of actionability across the population of comment parsings. On an individual comment parsing basis, a parsing from those subtopics had an 84%, 82%, or 74% chance of being actionable respectively. See Table 9 for text examples. For stakeholders in course redesign practice, an important next step would be to verify if generic feedback prompts, (e.g., Module 5 Feedback) elicit such high probabilities of feedback actionability with other audiences of students besides online instructors.

Interestingly, comment parsings grouped by categories and subcategories that were inversely related to Actionability (those with a negative log likelihood) had a much lower population likelihood of actionability ($e^{\beta} < 1$, p=0.01) and low individual comment parsing chances of being actionable (13% to 35%, p=0.01) than compared to categories positively associated with Actionability. In other words, parsings that were statements or had positive sentiment or topics regarding the module overall, the instructors' expertise, or other students were less likely to be actionable, but not to the degree that actionable comments tend to be negative. An important follow on study regarding the actionability of real-time positive feedback or real-time feedback on higher levels of course organization, (e.g., module, course overall) would be important to see if targeted feedback prompts could elicit higher levels of actionable feedback in these cases.

Table 15

Model 2: Fixed Effects: Two-Level, Dichotomous Outcome Generalized Linear Model, Actionable Parsing Grouped by Student Authorship with Predictors

Parameter	Intercept Log Odds	Std. E.	t	Sig.	Odds Ratio	Unit Probability	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	-0.2270	0.5840	-0.3880	0.6980	.	.	.	.
SubTopic=Support	6.2400	5.0600	12.3290	0.0000	512.859	99.8054%	190.0970	1387.3070
SubTopic=Accessibility	2.2130	0.4880	4.5300	0.0000	9.1431	90.1411%	3.5080	23.8180
SubTopic=Usability	1.6690	0.5120	3.3160	0.0010	5.3069	84.1442%	2.0020	14.9370
Purpose=Suggestion	1.5480	0.4610	3.361	0.0010	4.7021	82.4625%	1.9060	11.6030
SubTopic=Responsiveness	1.5260	0.5210	2.9280	0.0030	4.5997	82.1420%	1.6550	12.7680
SubTopic=Content Validity	0.9850	0.4000	2.4640	0.0140	2.6778	72.8099%	1.2230	5.8620
Topic=Module	-0.5780	1.3000	-4.431	0.0000	0.5610	35.9393%	0.4340	0.7250
Sentiment	-1.2380	0.0750	-16.394	0.0000	0.2900	22.4784%	0.2500	0.3360
SubTopic=Subj. Expertise	-1.5910	0.3180	-5.0080	0.0000	0.2037	16.9243%	0.1090	0.0380
Topic=Students	-1.6090	0.5800	-2.776	0.0060	0.2001	16.6727%	0.0640	0.6230
Purpose=Statement	-1.8660	0.3630	-5.138	0.0000	0.1547	13.4005%	0.0760	0.3150

The fair rater agreement level (Landis & Koch, 1977) between rater one and two for the subtopics categories, however, suggested that caution was needed with the subtopic results. Nevertheless, the Kappa score for subtopics in real-time feedback (Kappa=0.2430, p=0.000) was across 27 categories, about half of which had small frequencies. Agreement may be different for the significant predictors in Model 2.

Model 2 findings are important for stakeholders in course revision practice because they indicated that variance among per student averages in actionability likelihood was also significant for the topics and categories in Table 15. The results suggest the need for future research on which topics are actionable from student to student by student attributes. This would open the door for targeting the increased likelihood of comment actionability for specific topics or specific student attributes. We would also want to examine whether such an approach improves the likelihood of actionable feedback over the same feedback prompts across students or for generic ones.

Question 4 Results

Does real-time feedback result in more actionable student feedback than end-of-course feedback? As noted earlier, Nulty (2008) found online end-of-course response rates ranged from 23% to 47% with an average of 33%, which was much lower than the real-time feedback response rates observed in the current study (87%). Hence, we wanted to compare the real-time feedback to end-of-course feedback in the current study, not just in terms of response rates but also in terms of comment parsing actionability. Regarding response rates, we found a slightly higher rate (94%) among end-of-course responders if they just filled out ratings (See Table 16). Whereas, end-of-course responders who left end-of-course comments were much lower (43%) than the real-time feedback responders who left comments (87%). In either case, however, professional development students in this study (instructors and staff for online courses) were much more responsive with feedback than the online students in Nulty's (2008) review of end-of-course feedback rates.

Table 16

Student Counts and Percentages by Feedback Type Across 15 Course Sections

Students by Feedback Type	Student Count	Percent of Total Students
End-of-Course (EOC)		
EOC Feedback without Real-Time	22	7.97%
EOC Feedback	259	93.84%
EOC Feedback with Comment	119	43.12%
Real-Time (RT)		
RT Feedback without EOC	17	6.16%
RT Feedback without EOC Comment	147	53.26%
RT Feedback	253	91.67%
RT Feedback with Comment	240	86.96%
Total Students	276	100.00%

In terms of end-of-course actionability, we anticipated the end-of-course feedback comments, like comment feedback rates, would also have a lower percentage of actionable comments than real-time feedback. Therefore, we compared the students who gave comments in both end-of-course and real-time settings with a *Z*-test of the proportion actionable comments. To our surprise, however, we found that end-of-course feedback had a significantly higher rate of actionable comments (0.28, p=0.0) vs. real-time feedback (0.19, p.0.0) among students who left both real-time and end-of-course comments (See Table 17). In accelerated courses where real-time feedback is prompted throughout the course, Actionability of comments appeared to not be affected negatively by time for up to 2 weeks away from the targeted experiences or activity. Our concern that the length of the 2-week long class would affect the testing of recall-effects in real-time feedback may have been sported. Nevertheless, the total amount of actionable feedbacks was higher in the real-time feedback because it was prompted 9 times vs. the single prompting at the end of course. If one were seeking volume and variety of actionable course feedback, real-time feedback clearly would be the way to go.

Table 17

Z Test to Compare Proportions of Actionable Feedback Parsings Between Students' End-of-Course and Real-Time Feedback

Parameter	Proportion	Sample Size	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Real-Time	0.1901	1415	2.9	0.0041	0.1395	0.2405
End-of-Course	0.2802	232	.	.	0.2567	0.3035
Difference	.	.	.	.	0.0287	0.1515

Discussion

As we set out to examine the nature of real-time feedback from students to instructors we evaluated arguments about student feedback from the literature and proposed others of our own. Given the results of the current study we wanted to examine which premises and theories were supported.

Overcoming Disadvantages of Traditional End-of-Course Feedback for Course Redesign

End-of-course survey data are mostly ratings-focused (Fluit et al., 2010; Gravestock & Gregor-Greenleaf, 2008; Stieger & Burger, 2010), and do not underscore clear actions to take (Biggs & Tang, 2011; Bowden & Marton, 2003; Chapple & Murphy, 1996; Hendry et al., 2001; Saffran et al., 1994). The real-time feedback observed in this study, however, was comment-focused. Table 7 shows that about 87% of 276 students across 15 course sections gave at least 1 feedback with a comment. The 2,092 feedbacks were broken up into 2,970 feedback parsings based on when the topic changed in the comment. About 17% of the overall feedback parsings were actionable. Again, actionability is a dynamic affordance of the course revision practice carried out by instructor or instructional designers. We theorized that actionable comments facilitate the ease of flow in the course revision process by giving sufficient context that an action could be taken or design decision could be made regarding the course. Sufficient context would result by indicating the activity, when it happened, the student's sentiment, the issue at hand, and why the activity was affected. Certainly, the actionable, real-time comments in this study would provide an advantage to making course design decisions over end-of-course ratings feedback. The extent to which instructors and designers make course changes based on the feedback would be a strong future study.

Various researchers also suggested that end-of-course rating feedback is collected at the expense of qualitative data that could lend itself to knowing what changes to make (Boerboom et al., 2015; Hendry et al., 2001; Woloschuk, Coderre, Wright, & McLaughlin, 2011). The proposition was not tested, nor confirmed in this study. Even so, just because feedback is qualitative does not make it good feedback for design decisions. In this study, we observed that some types of categories of feedback were related to actionability more than others. Support, accessibility, usability, instructor responsiveness, and content validity comment topics were much more likely to be actionable than not, while the inverse was true for comments about the course module overall, the instructor's subject matter expertise, and fellow students. Furthermore, we found a strong inverse relationship between the comment sentiment and its likelihood of actionability, (i.e., positive comments were not actionable). Lastly, the type of communication also had a significant impact on whether the comment was actionable. Requests, Suggestions, Questions, and Statements ranged from highly actionable to highly not actionable in that order. Though the results should be viewed with caution due to the fair to moderate agreement in category classification, some practical steps to improving student feedback may be considered. For example, instructors or instructional designers may want to vary how they prompt students by topic or purpose of communication, (e.g., request or suggesting vs. statement) in order to affect the actionability of the resulting comments. Testing whether feedback prompts targeted at such themes in fact result in more actionable feedback in those areas would be a valuable contribution to the nascent, real-time feedback literature.

Other disadvantages of using end-of-course feedback for making course design decisions claimed in the literature included collecting it at the end of the course (Gravestock & Gregor-Greenleaf, 2008; Spooren et al., 2013), which has been related to more general comments (Nasser & Fresko, 2002). We found to the contrary, however, that in terms of percentage, comments were more likely to be actionable at the end-of-course (0.28, p=0.0) than in real-time (0.19, p.0.0) among students who left end-of-course comments. We had proposed that recall biases (Estelami, 2015; Nicol & Macfarlane-Dick, 2006) would likely be a contributing factor as to why end-of-course feedback would be less actionable than real-time feedback. Seeing that end-of-course feedback was more actionable, recall biases may not have influenced feedback actionability when collected within a 2-week window.

To the contrary, one reason why the end-of-course feedback had a higher percentage of actionability may have been that the real-time feedback primed the students for giving actionable feedback at the end of the course. Furthermore, Michael Cole's (1996) concept of prolepsis, "representing the future in the present" (Oakley, 2011, p. 283) would suggest that the students' expectation of the end-of-course survey may have affected how they carried out real-time feedback and vice versa. For example, it is possible that students could have figured they would have the opportunity to leave specific feedback at the end of the course rather than when prompted. Conversely, we saw a prolepsis effect on real-time feedback. For example, many of the feedbacks from Module 9, the last module of the course, were treated as an end-of-course feedback, such as, "Very good. Thanks for the good course!" or "Highly rewarding overall experience." These alternate explanations are worthy of further investigation.

The surprising actionability result in end-of-course feedback, however, should be viewed with caution. The 2-week length of the observed course and its ideal feedback-giving, student audience of instructors and staff for online learning may have contributed to the surprising result. Moreover, the 119 out of 276 students who left both types of feedback possibly introduced a self-selection bias. Additionally, these students' comments resulted in 1,415 comment parts across 9 observations of real-time feedback, while only 232 comment parts from a single observation resulted from end-of-course feedback. The substantial imbalance in the data for comparison may also have played a role in end-of-course feedback resulting in a higher percentage of actionable feedback in this 2-week course.

Actionability as a Dynamic Affordance

In this paper we answered Gašević et al.'s (2015) call for learning theory-based research by proposing dynamic affordance (Cook & Brown, 1999) as a way to define the actionability of student feedback as an affordance of data for decision-making. Dynamic affordance helps define the actionability of data in terms of its quality and its use in practice. We found, in fact that capturing feedback in real-time throughout the course and classifying the student comments into actionable, topic, sentiment, and purpose categories enabled the plausible identification of actionable feedback for course revision practice. This is important not only for helping reflective practice (Winchester and Winchester, 2014) among instructors, but also because many students are taught by instructors who do not participate in revising the course activities and content, where such decisions are made by content owners or instructional designers removed from the teaching of the course. Additional studies will be needed to determine how much stakeholders in course redesign practice make actual use of such feedback and what effect those changes have on students.

Though we addressed dynamic affordance in terms of course revision practice, for future research we also propose its application to how students give feedback. Cook and Brown (1999) claimed that the activity of knowing is epistemically distinct from knowledge and the former only exists through the enactment of one's practice, (e.g., riding the bike, carrying out the teaching, etc.). As Cook and Brown (1999) described, "for human groups, the source of new knowledge and knowing lies in the use of knowledge as a tool of knowing within situated interaction with the social and physical world" (p. 54). Knowing draws upon static knowledge, (e.g., tacit-individual, tacit-group, explicit-individual, and explicit-group) through Dewey's notion of productive inquiry (Hickman, 1990) where meaning is, as Cook and Brown (1999) would say, a "generative dance" of applying possessed knowledge as tools to construct new knowledge and knowing through active engagement with the world (See Figure 1).

Viewing the construction of new knowledge as co-creation between situation and activity not only ties knowing and knowledge to students' specific experiences in context, but also suggests that authentic feedback and reflection on those activities would also be highly situated. Furthermore, the tight coupling of situation and knowing suggests that feedback on learning activity collected outside of its context would be suspect to bias, affecting the quality of feedback data. Though we found in these initial findings that end-of-course feedback was more actionable in a short, 2-week course, we suspect that feedback collected at end-of-course for longer courses may be ineffective for activity design decisions, not just because of recall effects (Estelami, 2015; Nicol & Macfarlane-Dick, 2006; Woloschuk et al., 2011), but also because the unit of analysis is the course or instructor overall, whereas design changes to the course are at the activity- and content-level of the course. We suggest that it may be more difficult for students to draw upon static knowledge of activity minutia long after its occurrence, than to reflect on the knowing generated through the enactment of practical activity.

Limitations and Other Future Research

This study had significant limitations that researchers should address in future research on real-time student feedback. First, parsing sentences into comment parts based on topic change and then classifying a wide variety of comment topics into categories and subcategories are challenging methods to carry out consistently through manual efforts. For example, only, Actionability, Purpose, Topic and Sentiment had interrater reliability rates above 0.41 (moderate agreement level) in the 10% sample of second rater ratings done in this study. The rater agreement results indicated that model effects may have been over or underestimated in Question 3. Furthermore, the amount of manual effort involved in parsing text and text classification is substantial. Future researchers examining student feedback may want to use natural language processing and machine learning techniques (Martin & Jurafsky, 2000; Nastase, Koeszegi, & Szpakowicz, 2007). Though creating training sets for supervised machine learning models is also a considerable effort, such methods are consistently repeatable and explicitly sharable.

Second, the 2-week long course and its students (online instructors and staff) are not the usual setting and audience of most higher education courses. It remains to be seen if similar results to this study would be reproduced in 8- or 14-week courses that most undergraduate students encounter.

Nevertheless, this study was meant to be a proof of concept for real-time student feedback, inviting future research, rather than a definitive answer to what student comment types are actionable. Real-time feedback, as a method and area of inquiry, opens the door for several research directions. Of note is the finding that feedback authorship accounts for variability in the comment's actionability. Hence, targeting feedback prompts to student attributes and other predictors of actionability may help achieve high actionability and high comment diversity simultaneously (See Figure 4). Instructors and instructional designers may be able to vary feedback prompts across groups of students by student attributes or by topics that are likely to elicit actionable responses, rather than ask them all the same feedback question.

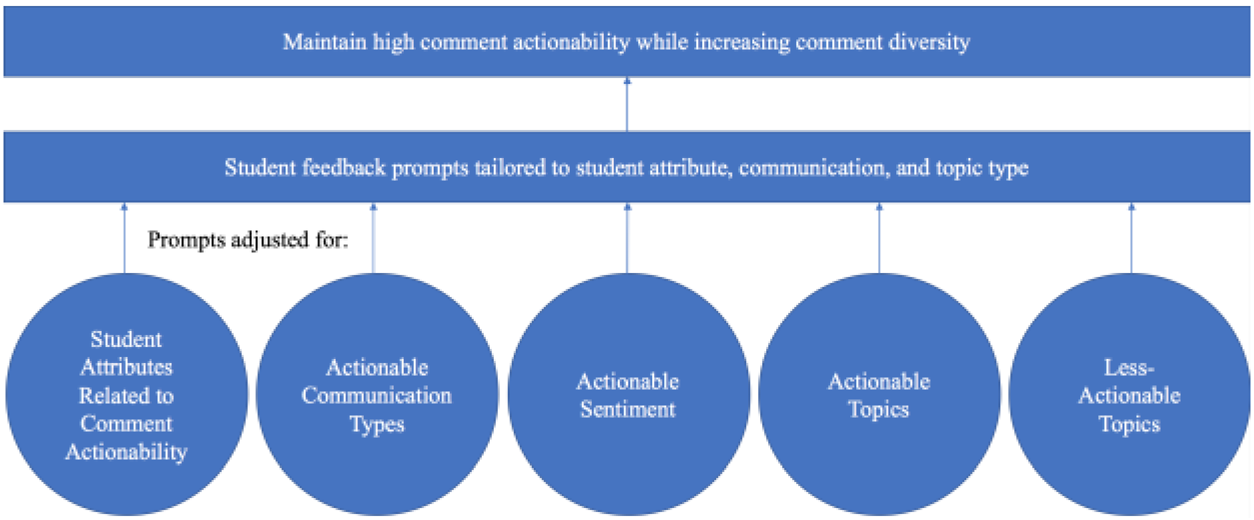


Figure 4. Adaptive student feedback capture

Furthermore, we would like to know if high, real-time feedback response rates persist across a variety of course formats, lengths and topics. And what factors predict real-time feedback rates. Other important questions to answer are included in Table 18.

Table 18

Future Research Questions on Real-Time Feedback and Actionability

Inquiry Area	Research Question
Increasing Actionability	Does prompting feedback on topics that are known to result in actionable feedback result in more actionable feedback?
	Does prompting feedback in new ways on topics that are known to not result in actionable feedback result in more actionable feedback?
	Does prompting feedback by focused communication purpose types, (e.g., suggestions, questions, etc.) result in more actionable feedback?
	Does presence of ratings-based feedback in courses reduce the likelihood of actionable feedback in comment answers?
Use of Actionable Feedback	What do instructional designers and instructors do with actionable feedback?
	What properties of feedback result in an instructional designer or instructor acting on that feedback?
Actionable Feedback and Course Outcomes	Does making course revisions on high rates of actionable feedback result in higher course outcomes more so than courses with low rates of actionable feedback?
Real-Time Feedback Time Effects	What affect do recall-biases, including the peak-end rule have on real-time feedback?
	Does real-time feedback prompted throughout the course prime students to give more effective end-of-course feedback?
Situated Feedback	Are real-time feedback topic variety and frequencies measures of situated feedback?
Real-time Feedback and student learning	Is real-time feedback from students an effective measure of learning?
Actionable in Aggregate	Are many individual comments about a topic that are not actionable on their own actionable in aggregate?

Conclusions

Real-time-collected student feedback appears to be an effective means of capturing actionable student comments in terms of course design practice and has significant implications for how stakeholder in course quality revise courses. For example, we found that collecting feedback in real-time can result in high levels of feedback. The median response rate across 15 sections of students was 9 out of 9 feedbacks with over 70% of students submitting 9 or more feedbacks. Hence, situating feedback prompts at or near the time of activity throughout the course can result in high levels of student feedback engagement, and may even prime students for providing higher quality end-of-course feedback than usual. This is an important difference between real-time feedback practice and the implementation of traditional end-of-course surveys. Situated feedback prompts at a minimum appear to produce higher actionable feedback counts than end-of-course surveys. For this reason alone, we strongly recommend that instructors and instructional designers implement real-time feedback.

Second, we found that real-time feedback resulted in a wide variety of comment topics and subtopics and that the likelihood comment actionability was explained by student authorship, sentiment, the student's communication purpose, and certain comment subjects. These findings open several doors for future research on how feedback prompts can be tailored to student attributes to achieve higher frequencies of actionable comments among varied situations and desired feedback themes.

Lastly, we found that in an accelerated course, with real-time feedback prompts throughout, end-of-course feedback comments had a higher percentage of actionable parsings than the average across real-time feedback comments (only from students who gave both real-time and end-of-course feedback comments). This suggests that real-time feedback may be affected by prolepsis (Cole, 1996) or even serve a priming role throughout the course, potentially overcoming recall effects normally associated with end-of-course feedback (Estelami, 2015; D. J. Nicol & Macfarlane-Dick, 2006; Woloschuk et al., 2011). Additional research is needed to see if real-time feedback helps overcome recall effects in regular term courses.

Ultimately, if instructors and instructional designers are to adopt this analysis approach as part of course redesign practice, however, the manual content analysis burden must be reduced. Natural language processing and machine learning must be used to classify comments into categories in a consistent and scalable manner. Because computational linguistic approaches to content analysis often rely on probabilities derived from large training sets (examples and non-examples of the construct), as a next step we recommend future research also focus on developing narrow training sets of critical, prioritized categories, (e.g., actionability, course logistics, etc.). With such automation capabilities available today, we look forward to near-term advances in the utility of real-time feedback for course redesign practice.

References

Abrahams, M. B., & Friedman, C. P. (1996). Preclinical course-evaluation methods at US and Canadian medical schools. *Academic Medicine*, *71*(4), 371–4.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*(3), 431-441.

American Association of University Professors. (2016). *Higher education at a crossroads: The annual report on the economic status of the profession, 2015-16*. Retrieved from https://www.aaup.org/sites/default/files/2015-16EconomicStatusReport.pdf

Atkisson, M. A. (2012). MOOC tweet sentient levels and their predictors. Unpublished manuscript. Department of Instructional Psychology and Technology, Brigham Young University, Provo, UT.

Atkisson, M. A., & West, R. E. (2016) A review of formative student feedback practice and supporting tools. Unpublished manuscript. Department of Instructional Psychology and Technology, Brigham Young University, Provo, UT.

Aultman, L. P. (2006). An unexpected benefit of formative student evaluations. *College Teaching*, *54*(3), 251–285.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, *87*(4), 656-665.

Beaty, L. (1997). *Developing your teaching through reflective practice*. Birmingham: SEDA.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, *17*(1), 48–62.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (3rd edition). Poland: McGraw-Hill Education (UK), Open University Press.

Boerboom, T. B. B., Stalmeijer, R. E., Dolmans, D. H. J. M., & Jaarsma, D. A. D. C. (2015). How feedback can foster professional growth of teachers in the clinical workplace: A review of the literature. *Studies in Educational Evaluation, 46, 47-52*.

Bowden, J., & Marton, F. (2003). *The university of learning*. London: Psychology Press.

Carle, A. C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers & Education*, *53*(2), 429–435.

Chapple, M., & Murphy, R. (1996). The Nominal Group Technique: extending the evaluation of students' teaching and learning experiences. *Assessment & Evaluation in Higher Education*, *21*(2), 147-160.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.

Cook, S. D. ., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*(4), 381–400.

Dennen, V. P., & Bonk, C. J. (2007). We'll leave the light on for you: Keeping learners motivated in online courses. In *Online and distance learning: Concepts, methodologies, tools, and applications* (Vol. 2, pp. 704–715). IGI Global.

Desai, B. (2014). A novel use of Twitter to provide feedback and evaluations. *The Clinical Teacher*, *11*(2), 141–145.

Estelami, H. (2015). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing Education*, *37*(1), 54–64.

Fluit, C. R., M, G, Bolhuis, S., Grol, R., Laan, R., & Wensing, M. (2010). Assessing the quality of clinical teachers: A systematic review of content and quality of questionnaires for assessing clinical teachers. *Journal of General Internal Medicine*, *25*(12), 1337–1345.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends: Linking Research & Practice to Improve Learning*, *59*(1), 64–71.

Goldfarb, S., & Morrison, G. (2014). Continuous curricular feedback: A formative evaluation approach to curricular improvement. *Academic Medicine February 2014*, *89*(2), 264–269.

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Higher Education Quality Council of Ontario Toronto. Retrieved from http://www.celt.iastate.edu/pdfs-docs/classclimate/research/2008_gravestock_course_evaluations_research_models_publication.pdf

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1209.

Heck, R. H., & Thomas, S. L. (2008). *An introduction to multilevel modeling techniques: Second edition* (2nd ed.). New York, NY: Routledge.

Heck, R. H., Thomas, S., & Tabata, L. (2013). *Multilevel modeling of categorical outcomes using IBM SPSS*. New York, NY: Routledge.

Hendry, G. D., Cumming, R. G., Lyon, P. M., & Gordon, J. (2001). Student-centred course evaluation in a four-year, problem based medical programme: Issues in collection and management of feedback. *Assessment & Evaluation in Higher Education*, *26*(4), 327–339.

Hickey, D. T. (2015). A situative response to the conundrum of formative assessment. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 202–223.

Hickman, L. A. (1990). *John Dewey's pragmatic technology*. Indiana University Press.

Jahangiri, L., Mucciolo, T. W., Choi, M., & Spielman, A. I. (2008). Assessment of teaching effectiveness in U.S. dental schools and the value of triangulation. *Journal of Dental Education*, *72*(6), 707–718.

Jara, M., & Mellar, H. (2010). Quality enhancement for e-learning courses: The role of student feedback. *Computers & Education*, *54*(3), 709–714.

Kamata, A., Bauer, D. J., & Miyazaki, Y. (2008). Multilevel measurement modeling. In A. A. O'connell & D. B. Mccoach (Eds.), *Multilevel modeling of educational data* (pp. 345–388). Charlotte, NC: Information Age Publishing.

Kezar, A., & Maxey, D. (2014). Student outcomes assessment among the new non-tenure-track faculty majority. *Occasional Paper*, (21). 1-30.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lang, J. W., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, *35*(3), 187–205.

Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching and Learning*, *2001*(87), 25–32.

Maistre, K. L., & Weston, C. (1996). The priorities established among data sources when instructional designers revise written materials. *Educational Technology Research and Development*, *44*(1), 61–70.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*(5), 707-754.

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, *99*(4), 775-790.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187-1197.

Martin, J. H., & Jurafsky, D. (2000). Speech and language processing. *International Edition*. Retrieved from http://www.ulb.tu-darmstadt.de/tocs/203636384.pdf

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, *52*(11), 1218–1225.

McKone, K. E. (1999). Analysis of student feedback improves instructor effectiveness. *Journal of Management Education*, *23*(4), 396–415.

Menges, R. J., & Brinko, K. T. (1986). Effects of Student Evaluation Feedback: A Meta-Analysis of Higher Education Research. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA: American Educational Research Association. Retrieved from http://eric.ed.gov/?id=ED270408

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, *27*(2), 187–198.

Nastase, V., Koeszegi, S., & Szpakowicz, S. (2007). Content analysis through the machine learning mill. *Group Decision and Negotiation*, *16*(4), 335–346.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301–314.

Oakley, T. (2011). Attention & rhetoric: Prolepsis and the problem of meaning. In C. Meyer & F. Girke (Eds.), *The rhetorical emergence of culture*. Berghahn Books.

Ramsden, P. (2003). *Learning to teach in higher education*. London & New York, NY: Routledge.

Ravelli, B. (2000). Anonymous online teaching assessments: Preliminary findings. Presented at the Annual National Conference of the American Association for Higher Education, Charlotte, NC. Retrieved from http://eric.ed.gov/?id=ED445069

Saffran, M., Conran, P. B., & Lacher, D. A. (1994). An approach to student-to-faculty feedback in the basic sciences. *Academic Medicine*, *69*(5), 413.

Satorra, A., & Muthen, B. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316.

Seldin, P. (1989). Using student feedback to improve teaching. *New Directions for Teaching and Learning*, *1989*(37), 89–97.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching the state of the art. *Review of Educational Research*, *83*(4), 598–642.

Stevens, J. J., & Aleamoni, L. M. (1984). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science*, *13*(4), 285–304.

Stieger, S., & Burger, C. (2010). Let's go formative: continuous student ratings with Web 2.0 application Twitter. *Cyberpsychology, Behavior, and Social Networking*, *13*(2), 163–167.

Thomas, S., Chie, Q. T., Abraham, M., Raj, S. J., & Beh, L.-S. (2014). A qualitative review of literature on peer review of teaching in higher education: An application of the SWOT framework. *Review of Educational Research*, *84*(1), 112–159.

US Department of Education. (1991). *Assessing teaching performance* (The Department Chair: A Newsletter for Academic Administrators No. 2(3), 2).

Wagner, J. P., Tillou, A., Nguyen, D. K., Agopian, V. G., Hiatt, J. R., & Chen, D. C. (2015). A real-time mobile web-based module promotes bidirectional feedback and improves evaluations of the surgery clerkship. *The American Journal of Surgery*, *209*(1), 101–106.

Weston, C., Le Maistre, C., Mcalpine, L., & Bordonaro, T. (1997). The influence of participants in formative evaluation on the improvement of learning from written instructional materials. *Instructional Science*, *25*(5), 369–386.

Wilson, J. H., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching*, *62*(1), 20–24.

Wilson, J. H., & Ryan, R. G. (2012). Formative teaching evaluations: is student input useful? In *Effective evaluation of teaching: A guide for faculty and administrators* (pp. 22–29). Society for the Teaching of Psychology. Retrieved from http://stp.wildapricot.org/Resources/Documents/ebooks/evals2012.pdf#page=26

Winchester, T., & Winchester, M. (2010). Exploring the utility of using a VLE for weekly student evaluations of teaching. In *International Conference on Information Communication Technologies in Education (ICICTE), Corfu, Greece*.

Winchester, T., & Winchester, M. (2011). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development*, *16*(2), 119–131.

Winchester, T., & Winchester, M. (2012). If you build it will they come? Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *37*(6), 671–682.

Winchester, T., & Winchester, M. (2014). A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology*, *45*(1), 112–124.

Woloschuk, W., Coderre, S., Wright, B., & McLaughlin, K. (2011). What factors affect students' overall ratings of a course? *Academic Medicine*, *86*(5), 640–643.

Dissertation Conclusion

One day at Stanford Graduate School of Business (GSB) in 2013, where I worked as an instructional designer, Dr. Baba Shiv, a distinguished marketing professor at the school with whom I had met only once months earlier, sent me an invitation out of the blue to meet a former student of his, Karan Chaudhry. Karan, a GSB alumnus, was the founder of DropThought, a Stanford StartX startup. He was in need of an instructional designer. That day, I joined DropThought, not just for instructional design, but also because I was taken aback at the serendipity between the company's vision for text analytics on customer feedback captured in real-time and my Ph.D. qualification project at Brigham Young University (BYU) on sentiment analysis of tweets in Massively Open Online Courses that I had recently completed. Soon after joining DropThought, I pitched to Karan a minimal viable product and the creation of a new real-time feedback division at DropThought that focused on student feedback. Karan and the DropThought Board agreed, as long as I could find a quick win in a pilot showing that students would use DropThought in a school setting. A dream of mine had just come true, the opportunity to design and bring to market an educational technology product that could benefit students on a broad scale.

Shortly thereafter, Dr. Taylor Halverson and Dr. Charles Graham were at Stanford for a few days on BYU business. Dr. Halverson arranged a lunch so I could ask Dr. Graham to pilot the yet-to-be built DropThought product for real-time student feedback in some of his classes. Dr. Graham agreed, and then Karan gave the okay to build a prototype with the DropThought team. The pilot was successful with the seven BYU education classes and became the source of the article in *International Review of Research on Open and Distance Learning*, *Exploring* *intensive longitudinal measures of student engagement in blended learning,* by Henrie, Bodily, Manwaring, and Graham (2015).

Because of the high levels of BYU student engagement with the prototype, Karan gave me the okay to design a full Software as a Service (SaaS) version 1.0 of DropThought for Education. The BYU pilot results and literature contributions also impressed Dr. Deb Adair of Quality Matters (QM) sufficiently that she agreed to have QM partner with DropThought in course quality assurance for itself and its member institutions. Dr. Adair was gracious enough to let me use the QM pilot data for this dissertation.

Now, almost 3 years later, DropThought has been acquired and I no longer am affiliated with the company, but DropThought Education lives on. When I left 15 months ago, about 400 institutions had instructors that used DropThought to capture student feedback in real-time.

I tell this story to recognize the important roles that many have played across Stanford, BYU, DropThought, Quality Matters, and Indiana University (my alma mater) to bring real-time feedback to higher education, without which, these contributions to the real-time feedback literature would not exist.

The purpose of this dissertation was to describe the growing trend of formative feedback in higher education, given the rise of technologies that enable feedback from students to instructors to be captured in real-time (Atkisson & West, n.d.). Surveys are well known and widely used in higher education, particularly end-of-course surveys. Despite near ubiquity, there has been no evidence that links end-of-course surveys to the improvement of instruction over time (Lang & Kersting, 2007; Marsh, 2007; Menges & Brinko, 1986; Wilson & Ryan, 2012; Winchester & Winchester, 2014).

Nevertheless, feedback from students in education is generally viewed as an important part of the process for making updates or redesigning course. Furthermore, because of a growing interest in formative feedback (Aultman, 2006; Desai, 2014; Goldfarb & Morrison, 2014; Hendry et al., 2001; McKone, 1999; Ravelli, 2000; Wagner et al., 2015; Winchester & Winchester, 2010, 2011, 2012) and given the growing number of tools that support the real-time capture of student feedback, including my own, I set out to index the uses of formative and real-time student feedback and the tools that support them in Article 1.

Article 1 identified a new model, the Lifecycle of Student Feedback, to describe the steps that scholars recommend and the steps that instructors carry out to capture and make use of formative feedback. The framework identified gaps between recommended practice and observed practice among the reviewed articles, such as in analyzing data. No studies in the review adequately described how formative feedback should be analyzed in an effective or efficient manner, the biggest drawback and adoption obstacle to the formative feedback practice. We also used the framework to outline the affordances of the formative feedback tools and aspects of practice found in the literature, including how frequently formative feedback is prompted. Instructors captured formative feedback at different rates from one study to another, which ranged from ad hoc and weekly to a per assignment basis. Lastly, we used the framework to outline gaps in the tools' abilities to support the recommended student feedback practice. We recommended that formative feedback tool providers support all the steps of the lifecycle of student feedback, not just the capture of the feedback. Such steps not supported within the reviewed tool sets included, Defining your Inquiry, Developing Questions, Analyzing Data, and recognizing student concerns. Facilitating these feedback steps through software would encourage the use of feedback best practices by reifying them in a tool that supports their everyday use. The hope is that such standardization would lead to better feedback capture and benefit.

The purpose of Article 2 was to define actionable feedback, describe its nature in terms of predictors, and test to what extent real-time feedback capture may result in actionable feedback. Our method generated a lot of student feedback. The student response rate was 87% overall with 70% of students (N=276) responding on 9 out of 9 occasions or more. A wide variety of feedback topics (2,092 comments into 8 topic and 27 subtopic categories), resulted from the analysis. We used correlations and hierarchical models to identify affordances of the feedback. Various practical recommendations resulted, including the need to adapt the type of feedback prompt to student attributes to provide actionable comments in consistently high manner. For example, if instructors or instructional designers want actionable feedback from students, they should be prompted to offer more than just to make statements, (e.g., requests, suggestions, or ask questions).

Lastly, we found that the level of actionability in the end-of-course feedback was higher than the average of real-time feedbacks. This was unexpected, as we had thought the study might be a case for demonstrating recall effects (Estelami, 2014; Nicol & Macfarlane-Dick, 2006) in real-time feedback. It appears that, to the contrary, real-time feedback may have had a priming effect on the level of actionability of comments in the end of course survey. Additional studies will need to determine whether real-time feedback priming exists or if priming is only an artifact of certain feedback audiences.

This research opens a new field of questions across many avenues. Is real-time feedback related to the improvement of instruction over time? In other student audiences, do the benefits of formative feedback in the literature pan out in real-time feedback, including increased volume and detail in feedback (Desai, 2014), greater reliability in ratings (Goldfarb & Morrison, 2014) and the opportunity to make changes to teaching (Ravelli, 2000; Winchester & Winchester, 2011a)? Furthermore, do students observe the cited benefits of formative feedback when real-time feedback is prompted in their courses including, timeliness of interventions (Aultman, 2006), increased student satisfaction from seeing instructor responsiveness (Aultman, 2006; Hendry et al., 2001; Ravelli, 2000), increased student engagement (Aultman, 2006), and increased ownership of learning (Aultman, 2006; Ravelli, 2000)? Additionally, it would be important to test whether following the recommend steps in the Lifecycle of Student Feedback proposed in Article 1 with a tool that meets the feature criteria would in fact increase the consistency of actionability of student feedback.

In conclusion, our initial findings on real-time feedback illustrate practical actions instructors can take to receive quality feedback from students. The opportunity is large for real-time feedback researchers to determine what impact the method ultimately has on improving conditions and outcomes for students.

Dissertation References

Atkisson, M., & Wiley, D. (2011). Learning analytics as interpretive practice: applying Westerman to educational intervention. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 117–121). New York, NY, USA: ACM. https://doi.org/10.1145/2090116.2090133

Aultman, L. P. (2006). An unexpected benefit of formative student evaluations. *College Teaching*, *54*(3), 251–285.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. *Cambridge Handbook of the Learning Sciences:* Retrieved from http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf

College Board. (2015, July 29). Average rates of growth of published charges by decade - Trends in higher education. Retrieved July 30, 2015, from http://trends.collegeboard.org/college-pricing/figures-tables/average-rates-growth-published-charges-decade

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development*, *21*(1), 5–21.

Desai, B. (2014). A novel use of Twitter to provide feedback and evaluations. *The Clinical Teacher*, *11*(2), 141–145.

Digest of Education Statistics. (2014). Retrieved July 30, 2015, from http://nces.ed.gov/programs/digest/d14/tables/dt14_326.20.asp

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, *38*(1), 188–230.

Estelami, H. (2014). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing Education*, *37(1), 54-64.* 273475314552324.

Goldfarb, S., & Morrison, G. (2014). Continuous curricular feedback: A formative evaluation approach to curricular improvement. *Academic Medicine February 2014*, *89*(2), 264–269. https://doi.org/10.1097/ACM.0000000000000103

Hendry, G. D., Cumming, R. G., Lyon, P. M., & Gordon, J. (2001). Student-centred course evaluation in a four-year, problem based medical programme: Issues in collection and management of feedback. *Assessment & Evaluation in Higher Education*, *26*(4), 327–339. https://doi.org/10.1080/02602930120063484

Hutchins, J. (1999). Retrospect and prospect in computer-based translation. *Machine Translation Summit VII, 13th-17th September*, 30–34.

International Educational Data Mining Society. (2015, July 29). Retrieved July 30, 2015, from http://www.educationaldatamining.org/

Lorin, J., (2014, October 8). Student debt. Retrieved July 30, 2015, from http://www.bloombergview.com/quicktake/student-debt

Lang, J. W., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, *35*(3), 187–205.

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, *99*(4), 775-790.

Martin, J. H., & Jurafsky, D. (2000). Speech and language processing. *International Edition*. Retrieved from http://www.ulb.tu-darmstadt.de/tocs/203636384.pdf

McKone, K. E. (1999). Analysis of student feedback improves instructor effectiveness. *Journal of Management Education*, *23*(4), 396–415.

Menges, R. J., & Brinko, K. T. (1986). Effects of Student Evaluation Feedback: A Meta-Analysis of Higher Education Research. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA: American Educational Research Association. Retrieved from http://eric.ed.gov/?id=ED270408

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Ravelli, B. (2000). Anonymous online teaching assessments: Preliminary findings. Presented at the Annual National Conference of the American Association for Higher Education, Charlotte, NC. Retrieved from http://eric.ed.gov/?id=ED445069

Rychlak, J. F. (1988). *The psychology of rigorous humanism* (2nd ed.). New York, NY: New York University Press.

SoLAR. (2015, July 29). Retrieved July 30, 2015, from http://solaresearch.org/

The NCES Fast Facts Tool. (2015, July 29). Retrieved July 30, 2015, from https://nces.ed.gov/fastfacts/display.asp?id=40

US News. (2015, July 29). Freshman retention rate. Retrieved July 30, 2015, from http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/national-universities/freshmen-least-most-likely-return

Wagner, J. P., Tillou, A., Nguyen, D. K., Agopian, V. G., Hiatt, J. R., & Chen, D. C. (2015). A real-time mobile web-based module promotes bidirectional feedback and improves evaluations of the surgery clerkship. *The American Journal of Surgery*, *209*(1), 101–106.

Wilson, J. H., & Ryan, R. G. (2012). Formative teaching evaluations: Is student input useful? In *Effective Evaluation of Teaching: A Guide for Faculty and Administrators* (pp. 22–29). Society for the Teaching of Psychology. Retrieved from http://stp.wildapricot.org/Resources/Documents/ebooks/evals2012.pdf#page=26

Winchester, T., & Winchester, M. (2010). Exploring the utility of using a VLE for weekly student evaluations of teaching. In *International Conference on Information Communication Technologies in Education (ICICTE), Corfu, Greece*. Retrieved from http://www.harper-adams.ac.uk/aspire/files/WinchesterTLWTdissertation.pdf

Winchester, T., & Winchester, M. (2011). Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development*, *16*(2), 119–131.

Winchester, T., & Winchester, M. (2012). If you build it will they come?; Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *37*(6), 671–682.

Winchester, T., & Winchester, M. (2014). A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology*, *45*(1), 112–124. https://doi.org/10.1111/bjet.12019