On Trust, Editorial Intent, and Recommender Systems for Supporting Higher Education

Taha Hassan

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science and Applications

D. Scott McCrickard, Chair

Kenneth R. Edmison, Jr.

Bart Knijnenburg

Sang Won Lee

Aug 14th, 2024

Blacksburg, Virginia

Keywords: Adoption, Trust, Power, Recommendation

On Trust, Editorial Intent, and Recommender Systems for Supporting Higher Education

Taha Hassan

(ABSTRACT)

Institutional support of higher teaching and learning *at scale* poses three unique challenges. The first challenge is poor institutional accounting of instructors' use of educational platforms and software, especially the learning management system (LMS). The second challenge is a deficit of trust among stakeholders with unique job roles, prerogatives, and editorial preferences. The third challenge is one-size-fits-all, open-loop, or stopgap support processes. To address these challenges, this three-phase dissertation project proposes a novel sociotechnical framework for institutional support using trustworthy educational recommender systems. This framework accounts for LMS platform contexts, multiple stakeholders, and editorial trust relationships. In its first phase, the project proposes "Depth of Use" (DOU): a first-principles framework of frequent LMS use-contexts. DOU is found to highlight low-adoption course cohorts, evaluate course design interventions, and improve IT emergency preparedness. The second phase of this project proposes a novel model of recommendation trustworthiness based in stakeholder allocation of RS editorial tasks. The study discovers a spectrum of faculty intentions about editorial division-of-labor and its frequent rationales, including student expertise, professional curriculum needs, authorship burdens at scale, and learner disengagement. In its third phase, the project investigates how editorial trust might be enhanced by transparency cues (guarantees, social proof, content tags). The dissertation concludes with a set of design guidelines to aid HCI practitioners in enhancing editorial transparency and algorithmic explainability, and increasing process efficacy of institutional support.

On Trust, Editorial Intent, and Recommender Systems for Supporting Higher Education

Taha Hassan

(GENERAL AUDIENCE ABSTRACT)

In higher education, supporting faculty effectively can be challenging, especially with technology use at scale. This dissertation reckons with three primary aspects of this challenge: inadequate tracking of how educators use learning platforms, low trust among different institutional stakeholders, and inefficient support processes. To address these challenges, we propose a novel framework to personalize instructional support using learning management systems (LMS) as platforms to reach out to faculty, interpret their technology needs, and deliver interventions using educational recommender systems (ERS). Our framework allows better understanding of faculty's LMS use, editorial intent, and trust of automation. It also highlights structural barriers to trust and process efficacy at universities. Finally, it delivers guidelines for the design of trustworthy educational recommendation and support processes.

For Amma, Baba, Sibtay, Ayma, Najaf, and Kumail.

Acknowledgments

Donald Scott McCrickard, my dissertation adviser extraordinaire.

Amma and Baba. My safety net.

Sibtay, Kumail, Ayma, Najaf, Azam, Manzoor, Masooma, Nazma, and Uzma. *Sangré*.

Bob, Daron, Matt, Larry, Lujean, Santhosh, Ognen, and Brian. My longtime collaborators and TLOS family.

Abdullah, Safdar, Adil, Sarang, Ammar, Satyajit, and Iqtidar. Longtime friends.

Lindah, Neelma, Derek, Morva, Tim, and Jaitun. My HCI family.

Disha and Sameeksha. My messy and talented Blacksburg family.

Madhura. My enduring source of support.

Contents

List of Figures

List of Tables

List of Abbreviations

DelMo Delegation Model

DoIT Division of Information Technology

DOU Depth of Use

E-Auth Editorial Authority

HYP Hypothesis

LMS Learning Management System

RQ Research Question

RS Recommender System

TEL Technology-Enhanced Learning

TLOS Technology-Enhanced Learning and Online Strategies

Chapter 1

Introduction

Institutions of higher learning frequently mandate the provision of instructional support to their educational technology organizations. This support comes in many forms, including instructional design, learning environment support, faculty professional development, software and hardware procurement, research grants, and technology evangelism. Supporting higher education at scale is a formidable challenge on the platform, stakeholder, and administrative fronts. On the platform front, while learning management systems (LMSs) offer globally popular native services for productivity, communication, and assessment [16, 60, 87, 106, 114, 145, 149], their capacity for actionable assessments of learning outcomes and faculty needs is presently limited [53]. On the stakeholder front, previous research [65, 109, 166, 167] notes that technology mistrust, cognitive burden of discovery, lack of technical support, budget limitations and career priorities can discourage EdTech adoption. On the IT administrative front [55, 66, 152], delivery of instructional support is challenged by communication deficits between faculty and institutional support personnel, minimal accounting of successful strategies for outreach and intervention, balkanization of expertise, risk-averse data governance, lengthy software acquisition processes, and staffing constraints. Several studies on technology-enhanced teaching and learning [14, 137, 157] have noted the disappointing collective outcomes of these challenges: tensions between faculty-initiated "grassroots" projects and administrative initiatives [70], course quality assessments and end-of-term evaluations fueling negative effects (absenteeism, economy grades, lower as-

`img/closed_loop.png`

Figure 1.1: A design vision of closed-loop, scalable institutional support rooted in (a) respecting faculty's LMS use needs and editorial preferences, (b) leveraging these needs to inform content delivery and personalization, and (c) evaluating faculty buy-in to microtarget specific support interventions (instructional design, professional development, tech evangelism, etc.)

sessment standards) [49], and general perceptions of dissolution of trust, process inefficacy, or lack of institutional recognition. We observe that all three of these challenges combine issues of **competence** (such as lack of LMS effectiveness for data analytics, or ill-defined administrative processes) and **value-alignment** (such as students' outcome bias, or faculty's priority stack and motivation for professional development), two fundamental aspects of stakeholder trust [117, 162]. According to previous literature on higher education management, addressing these trust barriers can improve the effectiveness of governance [160], scope of instructional support [7], teaching efficacy [118], student satisfaction [151], and alumni loyalty (recommendation, grants and donations) [25, 57, 146] for universities.

This dissertation argues that an effective solution to these trust barriers requires a closed-loop, scalable, and personalizable institutional support process (visualized in figure 1.1). This vision has two primary components. The first is an assessment of faculty's LMS use needs across the entirety of their preferred LMS apps and services, their trust of human and algorithm stakeholders, and their editorial intent. The second is the use of these evidence points to help curate personalized teaching and support content aboard the LMS. This support vision proposes to build the capacity for actionable LMS data analytics, reach a broad campus audience, and improve presently inefficient processes of software testing, faculty development and teaching support [137], addressing concerns of competence. Empowering the grassroots working relationships between institutional support personnel and course faculty by improving communication can also enhance value-alignment and overall trust. Coupled with automated content delivery, data-driven faculty relationship management can help IT administrators identify long-term faculty partners for initiatives targeting course redesigns, professional development, quality assurance, and curriculum management.

We therefore identify the research imperative of this dissertation project as to:

- evaluate **learning management system use** at scale, to support sensemaking of faculty needs and evaluate the efficacy of institutional support (Study I),
- understand the **editorial preferences** and **trust beliefs** of higher education stakeholders with unique work-based prerogatives and priority stacks, often arranged in teams and organizations (Study II)
- describe the implications of faculty's LMS use and editorial intent for the design of **educational recommender systems** to help curate teaching and support resources (Study III)

To accomplish these goals, this dissertation project proposes a full-stack study analyzing Virginia Tech faculty's (1) use of Canvas LMS, (2) editorial authority delegation and trust in teams, and (3) preferences of explainability, and transparency for a proposed LMS service: a content recommender system to assist with student learning, faculty professional development, and technology evangelism.

Three constituent studies were conducted for this project. Study I proposes *Depth of Use*

(DOU): an intuitive, first-principles framework to assess faculty's utilization of a learning management system. DOU is found to highlight low-adoption course cohorts, evaluate course design interventions, and improve departmental resource allocation. The details of this study are featured in our CSCW'24 [79] and ITiCSE'20 [77] papers, and reflections on DOU-informed redesign of a CS200 course at Virginia Tech appear in this book chapter (Springer LDT'22) [168]. Study II proposes a novel model of recommendation trust based in stakeholder allocation of RS editorial tasks. The study discovers a spectrum of faculty intentions about editorial division-of-labor, and its frequent rationales, including student expertise, professional curriculum needs, authorship burdens at scale, and learner disengagement (study details appear in our UMAP'21 paper [78]). Study III explores how editorial trust might be enhanced by transparency cues (social-proof, content tags, algorithm primers). The dissertation concludes with a synthesis of design guidelines to aid HCI practitioners and institutional support personnel in identifying editorial consensus, enhancing algorithm explainability, and facilitating reflection.

1.1 Background and Motivation

In this section, we describe the trust challenges faced by institutional support personnel that motivate this dissertation project. We then detail how learning management systems (LMS) and educational recommender systems (RS4TEL) can come together to address concerns of process inefficacy and technological barriers at the heart of these trust challenges, and vastly improve the effectiveness of support services.

Table 1.1: Institutional support in higher education: work roles, tasks, objectives and technology use. Work roles are designated by **ID** (instructional designer), **DEV** (developer), **PROJM** (project manager), **COMM** (communications), **FAC** (faculty), and **LEAD** (leadership). Primary work role corresponding to each task is highlighted in bold.

Work Roles	Impact Area, Work Tasks	Design and Research Objectives	Technology Use Cases, Strategies
ID, LEAD, DEV	Course redesign, retooling, program certification	Developing or evaluating a course, course website, educational technology, or learning environment	Learning data analytics, accessibility active learning, content creation, productivity, curriculum management, virtual environments, assessments, teaching evaluation
LEAD, PROJM, DEV	Support allocation, management	Designing the balance of institutional investment in support vs. automation: procurement, staffing	4Help, consultation, pandemic response, strategy: personnel vs. automation (teaching assistance, professional development, instructional design, AI)
COMM, PROJM, LEAD	Faculty outreach, tech evangelism	Developing or evaluating faculty outreach and communication strategies	Email campaigns, brown-bag talks, consults, demos, Canvas announcements, social media campaigns, AI-driven recommendations
PROJM, ID, DEV	Program management	Managing composite services and programs	Learning communities, networked learning, accessibility, inclusivity, design cohorts

1.1.1 Supporting Higher Education: The Work Practice

Institutional support at universities includes a broad array of services made available to faculty, including instructional design and retooling, learning communities, digital skills coursework, software testing and support, research grants, and new technology evangelism (table 1.1). These services are offered by a university's educational technology organization, and supported by other organizations including information technology, academic affairs, libraries, and administrative offices of colleges and departments. Previous literature on institutional management and governance at universities [17][165][23] identifies a notably broad agenda of work activities such as executive management, general administration, fiscal operations, academic affairs, and community relations. Our use of the term 'institutional support' refers only to institutional efforts for technology-based support of teaching and learning.

Through a pilot survey aimed at Virginia Tech Division of IT and TLOS staff in 2024, we sought to understand the broader work practice of institutional support, to highlight the broad space of UX research and design opportunities in the domain, and to acquire an up-to-date picture of institutional support strategies in the aftermath of strategic pivots and reorganizations at TLOS. Using Hartson and Pyla's UX evaluation framework [69], the survey focused on work roles, impact areas, challenges, utility of novel technologies, and personalization of support. 11 staff members across four teams (learning systems, support services, instructional design, learning services) responded. Our findings are summarized in table 1.1, and visualized in figure 1.2. Survey questions are presented in Appendix D.

The survey reveals four broad areas of work practice for institutional support: **course redesign and retooling**, **support allocation and management**, **outreach and technology evangelism**, and **program management**, and five primary work roles (instructional designer, developer, project manager, communications, leadership) working together to fulfill these work duties. We also asked our survey respondents to describe key barriers to trust (competence, value-alignment) between stakeholders [117], which are noted in red in figure 1.2. On the *technology competence* front, TLOS staff complain about anecdotal LMS use reporting and difficulty to import legacy course materials. On the *process competence* front, many note a lack of central data governance and poor accounting of frequent technology-related IT complaints as core challenges to trust. Finally, on the *value-alignment* front, TLOS staff point to faculty's research priorities, their view of compliance standards in course development as tedious and unnecessary, and their lack of support for instructional design programs deemed unpopular with department colleagues.

Lastly, we inquired our survey participants about their perceived overall utility of the following technologies and artifacts in addressing these challenges:

(a) **LMS use analytics** (understanding how thoroughly faculty uses Canvas services) img/wp_with_barriers.png

Figure 1.2: The work practice of institutional support visualized: key roles, impact areas, and barriers to trust (indicated in red).

(b) **Course personas** (facilitating conversations between IT and department leadership using archetypes of successful courses, and best practices)

(c) **Recommender system** for Canvas (a delivery system for personalized micro-learning, new tech demos, announcements, goals, assessments)

(d) **Faculty relationship management** (understanding which faculty members communicate with TLOS frequently, and why, including their response rates, competencies, needs, communication preferences, common tickets)

(e) **Decision analytics**: Applying data analytics (causal inference, what-ifs) to infer optimal strategies for hiring, staffing, material and human resource allocation in teams

Figure 1.3 reveals that 74% of our survey respondents perceived the utility of LMS use analytics (74%) as the highest for instructional design and course retooling, 86% considered faculty relationship management as useful in helping evangelizing new technologies on campus, and 79% described decision analytics (79%) as useful for support allocation and management. For new tech evangelism, recommender systems were perceived as most useful (80.5%) next only to faculty relationship management (86%). TLOS staff thus acknowledges the usefulness of two key components of our design vision (figure 1.1): LMS use analytics and recommendation for TEL, towards institutional support efficacy. We outline the primary research questions of this dissertation as they pertain to theses technologies, as follows:

1.1.2 Learning Management Systems

Service-based learning management systems are, increasingly, the choice platform for productivity, communication and assessment at institutions of higher learning [43], and they are frequently employed to scale pedagogies across disciplines [142]. However, despite significant previous work in learning analytics [10, 83, 102] over the past decade, prior research notes img/tech_util.png

Figure 1.3: Perceived utility (% agreement) of artifacts and analyses to support instructional staff at Virginia Tech (IT, TLOS) for the impact areas of course retooling, new technology rollouts and evangelism, and support allocation and management.

the limited capacity of LMSs to evaluate the impact of student engagement, instructional support, and faculty development on student outcomes [53]. There are several reasons. One, the existing adoption studies rely near-exclusively on self-reported LMS use, and there is no consensus on how to model **frequent use-contexts** of LMS-hosted native and third-party educational apps. Two, **multiple stakeholders** [4][8] in the higher education domain (faculty, students, instructional designers, LMS administrators, department leadership) have overlapping, but by no means identical, requirements for data aggregation and reporting. Three, a multitude of data sources (app metadata, course site content, team drives, social media) provide challenging, ever-expanding volumes of data to analyze, with complex and varied institutional, and governmental rules for access, analysis, and reporting. This identifies the first of our primary research questions (1.3.1): how can we identify the frequent use-contexts of LMS apps, and their impact on learning outcomes?

1.1.3 Educational Recommender Systems

Educational recommender systems (ERS or RS4TEL) are important facilitators of teaching and learning practices [104]. Previous work in the area has revealed an impressive breadth of their use-cases, ranging from suggesting learning objects personalized using learners' browsing history and preferences [94], to suggesting learning paths (sequences of learning objects) within a course [34], to suggesting next-term coursework [54] and peers for group work [126]. It has noted their ability to successfully integrate with adaptive hypermedia [41], learning management systems [37] and MOOC platforms [91] in order to filter and personalize materials, and support a variety of learning contexts (formal, informal, mobile, self-paced) and domains (classroom, workplace). For the use-case of institutional support in higher education, an educational recommender system aboard an LMS is less opaque, reaches a broader audience, and allows more personalization and iteration of content policies compared with traditional needs surveys and email campaigns. However, the domain of higher education is noted for its risk-averse data use practices, administrative walled gardens, and editorial prerogatives of faculty and department leadership [148][78]. Therefore, in order for this capability to be trusted by faculty across departments, we need to better understand faculty's initial trust perceptions and editorial intent.

For instance, figure 1.4 illustrates how different stakeholders allocate recommendation authoring (seed) and editing (veto) powers. Almost all faculty (96%) allocate RS seed and veto powers for faculty, while only 52% and 44% allocate the same for teaching assistants and students, respectively. Almost all faculty (96%) do not support article vetoing by students, and all stakeholders agree on decreasing overall editorial power for faculty, teaching assistants and students, in that order. On the one hand, this underscores the importance of stakeholder privacy in the domain as course staff contemplates higher algorithmic agency. On the other hand, it points to the variety of stakeholder tasks, duties, roles, prerogatives img/seed_veto.png

Figure 1.4: Virginia Tech faculty, TA and students on allocating the ability to seed (left) and veto (right) the recommendations for a 'Suggested Readings' recommender system aboard Canvas. 44% of faculty in our survey assigned the 'seed' task to students, compared to only 4% of faculty who assigned the 'veto' task to students. See Section 4.3 for details.

and potential **biases** we need to account for in designing a reusable educational RS. Existing RS work on trust or group dynamics generally does not grapple with differences in tasks and prerogatives. Herein, we are motivated to ask our second and third primary research questions: how do we interpret stakeholder standards for, and inform the design of, trustworthy educational recommendation, given a potentially uneven distribution of editorial authority, such as between teachers and learners?

1.2 Thesis Statement

Scaling of institutional support in higher education requires (1) helping the support personnel assess faculty's LMS platform use and editorial intent, (2) creating a closed-loop process for recommending personalized TEL/PD content based on faculty needs.

img/proj_structure.png

Figure 1.5: Dissertation project structure and constituent study parameters.

1.3 Research Questions

In the previous section, we outlined three broad research questions (RQs) that inform three constituent inquiries (**Studies I**, **II and III**) of the overall dissertation, respectively. We describe these research questions, and their corresponding sub-RQs and hypotheses, in detail as follows. Figure 1.5 outlines the overall structure of the dissertation project, and design parameters for its constituent studies.

1.3.1 RQI: Understanding Faculty's Learning Management System Use

What is faculty's utilization of a learning management system (LMS) for a given course, especially as a function of course context?

$$DOU = f(\text{course context}) \tag{1.1}$$

1. **RQI-a**: $DOU = f(\textbf{modality})$
2. **RQI-b**: $DOU = f(\textbf{participation})$
3. **RQI-c**: $DOU = f(\textbf{logistics})$
4. **RQI-d**: $DOU = f(\textbf{outcomes})$

The hypotheses corresponding to **RQI** are as follows.

(**H1**) Undergraduate courses have higher DOUs relative to graduate DOUs,

(**H2**) STEM courses have higher DOUs relative to non-STEM courses,

(**H3**) Online-only courses have higher DOUs relative to face-to-face courses,

(**H4**) Third-party app use significantly affects DOU,

(**H5**) Course DOU is linked to the #students enrolled full-time in the course,

(**H6**) Course DOU is linked to pageviews for the LMS course website,

(**H7**) Course DOU is linked to the average GPA awarded in that course, and

(**H8**) Course DOU is linked to the DFW rate of that course.

(**H9**) Course DOU is significantly linked to the #teaching staff members for the course, and

(**H10**) Course DOU is significantly linked to the instructor's prior enrollment in on-demand coursework for professional development.

1.3.2 RQII: Understanding Faculty's Editorial Intent and Trust Perceptions for Educational Recommendation

How do faculty's editorial preferences (trust intentions) for faculty, teaching assistants, students, and RS algorithm vary as a function of trust beliefs and course contexts?

$$\text{Recommendation preferences} = f(DOU, \text{course contexts}) \tag{1.2}$$

1. **RQII-a**: E-Auth $= f((Stakeholdertrust))$
2. **RQII-b**: E-Auth $= f(\text{DOU})$
3. **RQII-c**: E-Auth $= f(\text{Course contexts})$

1.3.3 RQIII: Supporting Faculty's Delegation and Transparency Preferences

How do the trust beliefs and intentions of course faculty vary with the use of transparency affordances, and with course and platform contexts? How can we improve the process efficacy of institutional support with these affordances and contexts?

$$\text{RS trust beliefs and intentions} = f(\text{DOU, explanations, course contexts}) \tag{1.3}$$

1. **RQIII-a**: RS trust beliefs and intentions $= f(\text{explanations})$
2. **RQIII-b**: RS trust beliefs and intentions $= f(\text{DOU})$
3. **RQIII-c**: RS trust beliefs and intentions $= f(\text{course contexts})$

4. **RQIII-d**: Institutional support efficacy = f(DOU, trust intentions)

1.4 Dissertation Overview

The rest of this dissertation is organized as follows. Chapter 2 presents a review of related works. Chapter 3 talks about understanding instructors' frequent use-contexts of a learning management system. Chapter 4 details an inquiry into the notion of editorial authority and its relationship with trust perceptions for higher education stakeholders, and Chapter 5 examines the mechanisms to enhance editorial trust. Chapter 6 describes our conclusions and directions for future work.

Chapter 2

Literature Review

2.1 Learning Management System Use and Contexts

We review the present state of related work in the evaluation of learning management system (LMS) use as follows. This includes previous research in human factors of LMS adoption (subsection 2.1.1), LMS success as an information system (subsection 2.1.2), facilitating institutional support and management (subsection 2.1.3) and student intervention (subsection 2.1.4).

2.1.1 Human Factors

There is considerable prior work on qualitative grounds for LMS adoption, like teaching and learning efficiency, generational student expectations, and institutional expansion and consolidation [30, 31, 121, 166]. For course instructors, the basic predictors of the pace of LMS adoption are departmental affiliation (STEM vs. non-STEM, say) and course modality (online vs. face-to-face, say). West et al. [166] conducted semi-structured interviews with 30 college instructors over two semesters, about primary use cases, teaching efficacy and efficiency, and overall satisfaction with Blackboard LMS. The study identified so-called 'integration challenges': course instructors finding it difficult to integrate LMS services into their teaching practices. This notion of 'integration' was echoed by McGill and Klobas [116] for the case of student adoption of WebCT, whereby students with a more favorable view of the 'task-technology fit' of LMS services were more likely to have higher LMS utilization. The authors also noted that instructor norms (instructor's view of LMS usability, support staff availability, and access to training resources) affected student utilization of LMS services favorably. Following an institution-wide transition to Canvas LMS, Wilcox et al. [167] surveyed user perceptions on frequent modes of use and platform limitations for Canvas LMS. They identified a generation gap in expectations between students and course instructors, wherein the pervasive student use of the mobile LMS app rendered a subset of Canvas sites - designed by faculty members for the desktop - ineffective in navigation, flow and content organization. We move this research forward by conducting a large-scale study of staff needs (scale, interoperability, ubiquitous access) that have informed Canvas LMS adoption across Virginia Tech.

2.1.2 Information Systems

Likewise, an information systems (IS) perspective on LMS adoption has been thoroughly explored over the years [5, 6, 131]. A bulk of these studies apply and evaluate a canonical model of IS success first discussed by DeLone and McLean [38]. The model factorizes the individual and organizational success of an IS into quality (system, information, and service), use (utilization, intention of use) and net benefits (impact on overall satisfaction, and intention of use) [39]. Adeyinka and Mutula [6] conducted a university-wide study of IS success factors underlying WebCT adoption and operationalized LMS utilization using nature of use (mandatory or optional), frequency of use, access and availability. They found use and intention of use both to be strong correlates of WebCT success. Fathema et al. [51] evaluated TAM using survey data on faculty and student attitudes about Canvas LMS at two public universities. They discovered that system quality and user self-efficacy were strongly linked to system use and perceived usefulness. They also noted that system quality is a multi-faceted notion that incorporates issues like design aesthetics, flexibility of access, degree of customization, and multimedia support. Ngai et al. [124] reported a stronger effect of the perceived usefulness and ease-of-use on system use relative to that of attitude (interest expressed towards adopting a new system). These studies largely employ user-reported system use in their analyses. Nonetheless, there are some early instances of LMS use modeling such as Ozkan and Koseler [131], where study participants reported system use as the number of hours spent daily, on course-related activities with U-Link using a desktop or web application. We contribute a vendor-agnostic resource-specific metric of overall and service-level LMS use.

2.1.3 Institutional Support and Management

Review studies by [155], [171], and [109] note that an instructive, albeit limited body of educational research exists on institutional and department-level inquiries into faculty mentoring and professional development [19, 103], technical support [152], instructional consultations [55], online teaching policies [90, 108], faculty incentive schemes [138], and support systems for students [120]. The COVID-19 pandemic has brought renewed attention, for instance, to faculty needs [164] that hinder the development of online and distance education coursework. In a case study of obstacles to distance education [138], faculty and administrators cited the need for instructional design and technical support to aid digital skills development, and reduce the time overhead of course development. In [152], the authors found that technical difficulties in a digital skills training lowered the participants' test scores, and challenged their pre-training motivation. They recommended organizations invest in technical support, outreach to motivate potential learners, and interruption-free training environments. In comparison, a study by [90] examined frequent concerns voiced by administrators in selecting online technologies for post-secondary distance learning. Cost (time, money and manpower) of delivery and support, especially at scale, was found to be the biggest concern, followed by vendor lock-in, and inequities of technology access, especially for broadband internet, among students. The authors noted that a lack of adoption models appeared to diminish the administrators' confidence in open-source software, and smaller institutions were more likely to favor piloting open-source tools. They also found that institutions with stable enrollments were more likely to consider the effects of low-cost technology solutions on student perceptions, relative to institutions with student retention challenges.

This research emphasizes adoption as a key administrative return-on-investment (ROI) metric. It also notes that differences between stakeholder priorities and between technology needs of course cohorts are often revealed and tested at scale. Our contribution to this discussion is a vendor-agnostic adoption measurement and claim-testing strategy applicable to any number of courses, staff-favored apps and support strategies. We contend that DOU discourages balkanization of expertise and aids data analytic reuse with a tool for discussion and consensus-building among stakeholders.

2.1.4 Learning Analytics, Intervention, and Recommendation

A discussion of the key drivers of learning analytics research in Ferguson [53] and Dawson [35] notes how native LMS data analysis, visualization, and recommendation capabilities are presently limited and far from being widely adopted. A lot of student activity is external to the LMS, the data volume is huge and ever-expanding, and there is little standardization of the data aggregation and reporting methods, viz-a-viz critical use-cases for all stakeholders involved (faculty, students, instructional designers, LMS administrators, department leadership). These problems persist even as in the past two decades, inroads in educational data mining [46, 139, 140] have helped advance the state of the art in predictive modeling of student engagement, learning and achievement [18, 32, 33, 83]. Simultaneously, LMS log data analyses have been used extensively to model student and faculty use-contexts [27, 115], and to improve LMS features [52], often for specific disciplines and pedagogies [77]. Improving existing pedagogies, assessing learning outcomes and risk-of-failure for students [45, 80], and recommending materials to support learning and course design [3, 37, 104, 119] are formidable, well-studied research problems in the e-learning and higher education communities. One of the early instances of this approach is Course Signals at Purdue [10]. Course Signals uses students' course outcomes, frequency of interaction with the LMS (Blackboard Vista), prior academic history and demographic information to ascertain a failure-risk measurement. In [169], a short-term warning system for ailing students models the early-term drop in clickthrough rates for modules of an online course. [102] describe a similar early-warning system which identifies isolated students using an analysis of ego networks and micro-communities of high-ability students on an online course forum.

However, a coherent framework for sensemaking of large-scale, multi-service LMS data is essential to supporting adoption and trust [71] of these solutions at scale across courses, departments, and institutions. We build on the reviewed works in this section to describe a cohort analysis and policy claim testing approach. Instructional designers and department leadership can use DOU to test the efficacy of faculty development initiatives, course redesigns, teaching support, and LMS evangelism.

2.2 Modes of Editorial Trust in Recommendation

In this section, we review frequent trust domains and framings, existing models of delegation in human-AI and team-AI collaborative work, and design of explainable recommender systems. We demonstrate why a view of recommendation trust and transparency linked to editorial division of labor can further our understanding of organizational or work-based felt prerogatives, and help remedy associated obstacles to trust.

2.2.1 Trust Domains and Framings

Trust is an umbrella term for several distinct if connected problems across fields as diverse as recommender systems [159], information systems [117], user modeling [78], organizational psychology [11], and game theory [20]. Algorithmic and heuristic models of trust often inform interaction between human and software agents, for instance, in social and sparse information graphs and recommender systems [112][64][74]. The user-experiential notion of trust, in contrast, is typically operationalized as an individual or group attribute in relation to a collaborative web-based system [117][95][96]. Group recommender systems [61][9] model the consensus of independent recommendation users with distinct preferences and powers. A subset of this work relies on social choice theory to find rating fusion schemes (most popular items, least misery for group members, etc.) [24] while others [147] incorporate notions of power imbalance and overlapping behavioral tendencies to yield consensus strategies.

General models of trust using a sociological, psychological, or game-theoretic lens have also been well-studied over the years. These often model the risk assumed by a user in their *trust behaviors* in relation to a market or a digital artifact, and the beliefs, attitudes, and intentions that precede, inform, and moderate the perceived risk. McKnight [117] and Gefen [63] study the qualitative factorizations of trust perceptions, such as competence, benevolence, and integrity, for users of online legal advice and e-commerce websites. Mui et al. [122] propose a computational model using the notion of 'reciprocity' (the bi-directional exchange of favors or revenge) to infer social reputation and trust. Braynov and Sandholm [20] discover that a misrepresentation of agents' trustworthiness or distrust can results in sub-optimal degrees of social welfare, profits and cooperation in a bilateral negotiation game. Note how aforementioned models often capture trust exchange between stakeholders, enhancing or undermining their *reputation*, and by implication, trustworthiness of their social neighbors in the process. In organizational psychology, studies of workplace attitudes [11][21] examine trust as an antecedent cause, an outcome variable, a mediator, or a moderating influence.

2.2.2 Trust Beliefs, Intentions and Behaviors

An important qualitative factorization of trust is contributed by McKnight [117]. McKnight's meta-model posits that **trust beliefs** inform **trust intentions**, which in turn inform **trust behaviors**. This meta-model borrows its canonical notions (beliefs, intentions and behaviors) from the larger social-psychological framework of Theory of Reasoned Action (TRA) [56]. It further argues that **disposition** and **institution** are antecedents to trust beliefs. We describe these notions in a recommendation context as follows.

Trust beliefs Trust beliefs refer to the user-attributed values or standards for a recommender system. McKnight [117] observes that prior models of trust in information systems can be grouped along three key dimensions (**CBI**): competence, benevolence, and integrity. Competence is the ability of the trustee to fulfill the truster's needs and objectives. Benevolence is the trustee's inclination to care for and act in the trustee's interests. Integrity is the trustee's inclination to be honest and true to their contract.

Trust intentions Trust intentions signal a user's willingness to depend on the recommender system, for instance, by providing personal information, following advice, or making a purchase.

Trust behaviors Trust behaviors are user tasks performed (i.e. adoption or engagement) or risks taken (e.g, consumption and purchase behaviors) in response to information attained via the recommender system.

Disposition and Institution 'Disposition' and 'Institution' describe two antecedents of trust beliefs, according to McKnight [117]. These identify an individual's general tendency to trust people and institutions governing the recommender system (e.g., the web or an LMS), respectively.

We propose a multi-target view of an individual's trust perceptions towards domain stakeholders. In a recommendation context, this requires modeling of (a) the degree to which this individual believes each stakeholder possesses competence, benevolence and integrity overall (trust *beliefs*), and (b) the RS editorial tasks of sourcing, vetoing, rating, and commenting this individual assigns to all stakeholders (trust *intentions*). This model designates the recommendation algorithm as one of the stakeholders. Prior literature on *anthropomorphism* in computing observes that users of a technology artifact form beliefs about trusting the artifact that are similar in language and composition to their beliefs about trust of other humans [88]. Study II survey and interviews suggest the same for stakeholders in higher education (faculty, teaching assistants, students). Analyzing the trust perceptions of human and machine stakeholders together can help assess how both reinforce or diminish each other for different course and editorial contexts.

Why editorial authority? One, we find that there are distinct job-related *roles*, and an informal editorial process of *task allocation* in the domain of higher education. In our survey and interviews for Study II, we observe that stakeholders frequently refer to instructors' *prerogative* in authoring and maintaining the recommended readings. It appears that all stakeholders perceive distinct editorial roles and tasks implied by their job title, so it is useful that a conceptual model of trust should contend with these roles and how tasks are negotiated among them. We discover that these allocations are linked to trust relationships between the stakeholders.

Two, editorial intentions express the intent to trust. This is because trust intentions are typically understood as the willingness to assume some risk in interacting with a digital artifact. McKnight et al. note that the customer of an online retail system indicates their trust by volunteering personal information or making a purchase. In higher education, a faculty member can delegate recommendation authoring or editing tasks, with risks like mishandling of student feedback, misinformation and spam, or absenteeism.

2.2.3 Delegation and Human-AI Collaboration

Existing literature in human factors, general automation, and human-machine systems has long investigated the problem of allocating functions between humans and machines. Influential works in these fields have modeled this deference, or *delegation* to automation on a spectrum ranging from machine non-reliance to machine autonomy [133, 150], and examined its efficacy as a function of task structure, machine ability, human cognitive needs, control modalities, and trust [100, 127]. This research has journeyed into contemporary investigations of human interaction with complex autonomous systems, in particular issues of human-autonomy teaming (HAT), collaboration, task-sharing, and AI-assisted decision-making [22, 47, 50, 59, 132].

Lubars and Tan [101] report trust and choice difficulty as most significantly correlated with preference for optimal human-machine delegation across a diverse set of tasks (e.g., scheduling a business meeting, diagnosing flu, hiring a new employee, judging a defendant's recidivism risk, etc). Authors find trust beliefs of AI *competence* and *value-alignment* to be significant influences on delegation preferences, but not the need for AI interpretability, human motivation, and perceived risk. On the competence front, Fugener et al. [59]

note that human ability and expertise (such as lack of metaknowledge) can affect delegation efficacy and performance in image classification tasks. Similarly, Hemmer et al. [82] similarly observe that task performance and task satisfaction for image classification improve with AI-delegation, and both relationships are mediated by self-efficacy. However, as Helberger et al. [81] observe in a case study of journalistic AI, selecting values and evaluating alignment are complex challenges encoding domain-specific processes and stakeholder needs. For instance, professional values of editors (transparency, diversity, autonomy) [89], institutional editorial authority and process credibility [67], and consumer-centered values of choice effectiveness and satisfaction all interact in determining long-term value for news stakeholders. Technology-use values (for instance, privacy, sustainability, benevolence, integrity, etc.), and their interpretations and tradeoffs are often a function of the application domain [72, 92, 144]. The emerging enterprise of value-awareness in recommendation [36] highlights crucial perspectives for CSCW and HAT researchers on how to create long-term stakeholder value and strengthen organizational trust and credibility. Influenced by these works, we contribute **DelMo**, an AI-integrative, multistakeholder formulation of the delegation of editorial labor in higher education, expanding on the task list examined in Hassan et al. [78] (see Section 5.2.1 and Appendix D.2.2). This helps us systematically distinguish between degrees of preferred editorial delegation for a multitude of higher education stakeholders, and uncover the values influencing this consensus, especially in a power-asymmetric context like a teacher-learner relationship.

2.3 Trust and Explanations

Explanations are a transparency cue used to communicate the *intent* and *process* of the underlying recommendation algorithm [58], to *persuade* the user to participate (e.g. interact, share information, make a purchase) and increase their confidence in the recommender system [156], or to explain the *tradeoffs* worthy of consideration in decision-making [162]. Explanations are routinely used to enhance the perceived explainability and user trust of knowledge-based systems and decision support systems. Previous research on explanations has evaluated their utility for an impressive breadth of recommendation use-cases (including music, books, fashion, finance, health) and UX evaluation goals (transparency, trust, satisfaction, persuasion) [28][44][62]. Several taxonomies of RS explanations have been researched over the years [58][29][141]. A recent review of 85 foundational works in human-centered explainable AI [141] notes that effect of explanations on user trust is positive to mixed or indeterminate, subject often to explanation type, application domain, and evaluation strategies. There have been some recent investigations of trust and transparency in curation and recommendation of learning materials [129][128][78]. Ooge et al. [128] evaluate the effect of explanations on initial trust in an e-learning platform directed at adolescents. The authors report an increase in multidimensional trust with explanations, but note a mixed overall reception from users along with concerns of utility and degree of customization. In a related work, they observe that visualizing the impact of RS algorithm control can enhance trust perceptions of transparency regarding the e-learning platform [129].

It is not uncommon for contemporary XAI research to focus on supporting users' knowledge of system behavior [22][125][44]. Our research, in contrast, pursues their assessment of the curators' credibility, implied in the *human-human* and *human-RS* initial trust relationships, and the potential *barriers* to these relationships expressed in stakeholders' need for editorial transparency. Wang and Benbasat [162] examine common obstacles to trust in recommendation agents, for instance, *agency relationship* and *high choice discretion.* Agency relationships are characterized by *asymmetry of information* and *goal incongruence* between users and the algorithm (machine stakeholder). In our work [78], however, we observe that in a higher education context, this agency relationship is not always perceived as a trust obstacle by students, but a necessary feature of the prerogative of faculty as in-charge of all course content. In Study II, we report obstacles to editorial trust expressed by faculty (authorship burdens, risks of misinformation, student disengagement and disproportionate attention to outcomes). We hypothesize that these obstacles point to underlying deficits in perceptions of competence, benevolence and integrity among stakeholders, and explanations are a potential remedy for these deficits.

Authorship cues (AC) signal the *process* (source, target, and rationales) of editorial judgments about recommended course materials ("X% of faculty found this helpful for students in explaining merge sort"). They can also potentially provide material evidence of student participation, and by association, their competence and integrity in editorial matters, to course staff. A reference to RS consumption patterns of high-scoring subset of students in the social proof explanation is one way of harnessing the outcome bias obstacle to trust. Item rationales (IR) can help reduce information asymmetry between stakeholders and the RS algorithm by enhancing their knowledge about the sources and subject-matter relevance of recommended items. Algorithm attributes (AA) highlight the RS algorithm's capability to, for instance, automatically flag and remove malicious content can potentially alleviate the perceived misinformation risk in adopting a RS for high-enrollment courses.

2.4 Chapter Summary

We review literature on faculty needs driving the adoption of learning management systems and educational recommender systems. We discover that existing models of LMS use largely rely on qualitative self-reports, challenging their generalizability. We also discover that trust is a foundational value for the output of an educational recommender system, and the editorial dynamics in the higher education domain call for a thorough, respectful assessment of initial trust beliefs, knowledge curation prerogatives, and transparency expectations.

Chapter 3

Understanding Faculty's Learning Management System Use

Online learning management system (LMS) tools are, increasingly, the primary infrastructure for productivity, communication and assessment at institutions of higher learning. Measuring their utilization and impact is thus, critical to assessing the efficacy of online teaching and learning at scale. Existing models of LMS utilization, however, are largely qualitative, opaque to individual LMS tools, and difficult to generalize to the needs of multiple domain stakeholders. Study I proposes **DOU** ('depth of use'), a novel method for computation of intuitive ordinal rankings (low, medium, high) of LMS utilization using instructional design-driven taxonomies. DOU is based on a taxonomy of Canvas use motivated from a series of informal brainstorming sessions with administrators of a DoIT/TLOS course development program at Virginia Tech. The study proceeds to test hypotheses about the relationship between DOU rankings and course meta-attributes (modality, participation, logistics, outcomes), followed by expert reviews. On the whole, Study I seeks evidence for the efficacy of DOU in *interpreting* and *supporting* stakeholder decisions about adopting new LMS tools (faculty), allocating institutional support (IT administrators), informing LMS evangelism (instructional designers) and assessing administrative policy impact (department leadership).

img/jour_maindiag6.png

Figure 3.1: Study I overview (clockwise from top left): LMS use taxonomy development (Step 0), DOU estimation (Step 1), and hypotheses tests (Step 2). We estimate a Canvas use score (overall DOU) for each course in our analysis. We then test hypotheses about the relationship of DOU with course modality, participation, logistics, and outcomes. We also identify key DOU applications areas (adoption analytics, support allocation). See Section 3.3 for the datasets and methods utilized in each step.

3.1 Study Objectives

To answer its primary research question (section 1.3.1), Study I proposes a novel measurement model called 'Depth of Use' (DOU) [77]. This framework assigns an ordinal DOU score of LMS use (low, medium or high) to each of forty-thousand college courses offered between 2017 and 2019 at Virginia Tech. DOU uses a *vendor-agnostic* taxonomy of LMS use developed in collaboration with Virginia Tech instructional designers. Figure 3.1 describes the overall approach. We then hypothesis-test (ANOVA, multivariate regression analysis) these scores against course attributes like modality, participation, logistics and outcomes. This helps determine the frequent contexts where faculty and staff might deem a subset of LMS services effective. For instance, the study discovers a consistent impact of overall LMS use on learning outcomes, and increasing reliance by faculty on tools that favor scale, ubiquitous access and interoperability. Finally, we discuss three key applications of DOU, to a) help faculty members assess the relative utility of LMS services and legacy apps, b) aid instructional designers in measuring and improving the scope of interventions and LMS evangelism, and c) help LMS administrators identify the technology needs of actionable low-adoption cohorts. The framework forwards a *multistakeholder* view of LMS utilization, in that alongside learning analytics, it supports claim testing and cohort analysis for policy decision-making, an avenue with lesser treatment in educational research literature in the last decade [109].

3.2 Depth-of-Use: Background and Definitions

In Spring 2013, the Division of IT (TLOS) at Virginia Tech began offering a course development program for faculty interested in developing online, hybrid, and flipped classrooms [12]. This program was structured as a semester-length course with faculty-designer interest groups meeting weekly to design a new course, and building competency in topics including active learning, self-paced modules, flipped classrooms, lecture capture, accessibility, and copyright and fair use. Course faculty enrolled in the initiative worked on weekly assignments targeting syllabus review, assessments, online pedagogy, student and classroom management, and alignment with learning outcomes. As Canvas emerged at Virginia Tech [123] as the campus-wide LMS in 2015, TLOS commissioned Study I to evaluate the effectiveness of Canvas course sites emerging from the program. The first step of this study

Table 3.1: A taxonomy of LMS use-contexts.

LMS Resource	**Use Context**
Announcements (**An**)	**0**: None; **1**: Placeholder announcements; **2**: At least one per week or course instrument
Syllabus (**S**)	**0**: None; **1**: Syllabus under *Files*; **2**: File previewed/embedded under *Syllabus*
Discussions (**D**)	**0**: Discussions disabled; **1**: No discussion activity; **2**: Discussion groups with activity
Assignment Delivery ($\mathbf{A}_d$)	**0**: No assignments on LMS or placeholders; **1**: Link to DOC, ZIP or 3rd-party app; **2**: Assignments fully hosted on LMS
Quiz Delivery ($\mathbf{Q}_d$)	**0**: No assignments on LMS or placeholders; **1**: Link to DOC, ZIP or 3rd-party app; **2**: Quizzes fully hosted on LMS
Assignment Submission ($\mathbf{A}_s$)	**0**: No file upload, likely paper or 3rd-party app; **1**: LMS file upload; **2**: LMS text entry
Quiz Submission ($\mathbf{Q}_s$)	**0**: No online submission, likely paper or 3rd-party app; **1**: Submission within LMS
Gradebook (**G**)	**0**: No grading activity in LMS; **1**: Comprehensive grading for all assessments
Files (**F**)	**0**: No files; **1**: Course resources under *Files*

(figure 3.1) was a series of informal brainstorming sessions with the program administrators: three instructional designers, a director of learning experience design, and a director of IT software development, who expressed two preliminary research objectives:

- **Objective 3.A**: Understanding the degree to which faculty participants were applying the instructional best practices emerging from the program in their teaching practice,
- **Objective 3.B**: Understanding if faculty participants' engagement with the program was linked to an improvement in student engagement for their courses, relative to non-participants,

In service of these objectives, the administrators expressed significant interest in devising a course-level metric of user engagement with Canvas as a first step in creating a "performance img/dou_devs.png

Figure 3.2: Brainstorming session aids for Canvas depth-of-use development: three early, data-driven attempts to capture Canvas use from the page request log ((a)-(c)), and a preliminary model-driven summary of Canvas usage tiers ((d)).

scorecard" for the program (figure 3.2). Weekly and monthly page request counts from Canvas logs (normalized by course enrollment) were the initial metric of choice, followed by page request compositions and semester-length timelines. The author recorded notes on their laptop during the sessions. The administrators' discussions centered on whether the metric adequately highlighted the differences between cohort participants and non-participants, and could facilitate conversations between designers and TLOS's faculty clients - with a broad range of quantitative competencies - on how best to achieve this improvement in student engagement. Figure 3.2a through 3.2c illustrate the attempt to use (1) weekly course page request counts, (2) department scoreboards, (3) page request categories, and (4) semester-level page request time-series to highlight the distinctions between program participants and non-participants, and identify popular Canvas services and use-patterns.

The designers noted how these aids created difficulties for their faculty clients on several fronts:

- Defining a meaningful baseline LMS engagement is challenging with a continuous-valued metric ("*it is difficult to know if 50% Canvas site viewership is good enough*")
- Identifying the apps (LMS-native or third-party) most effective in teaching of a course is not possible with site-level metrics ("*overall site viewership does not tell us if the syllabus page we helped faculty design is working for them*"),
- Quick inference of best practices for faculty with low quantitative abilities is not made easy with tables and charts ("*not all faculty clients like numbers*")
- A variety of legacy tools are in use at different Virginia Tech colleges and departments, and site-level viewership is not reliable in comparing Canvas use across departments ("*one department's 50% Canvas use might be another department's 25% Canvas use*").

Based on this feedback, three key requirements emerged: the need for an **ordinal metric of LMS engagement** instead of a continuous-valued one, the focus on **taxonomizing faculty's Canvas use** instead of all LMS users, and the flexibility to incorporate **the use of new tools and services**. In the subsequent sessions, the research team therefore pivoted away from aforementioned page request visualizations (figure 3.2a-c), and towards a potential rule-based Canvas system of use applicable across departments. Figure 3.2d showcases an early Canvas use taxonomy using three services (announcements, discussions, assignments).

In the final brainstorming session, we solicited a simple tabulation, or taxonomy (table 3.1), of the designers' mental models of Canvas use, which we evaluated for consensus, consistency, and accuracy. We used the final two questions listed in Appendix E.1 to create an aggregate list of (a) how all participants taxonomized the use of LMS services, and (b) how they preferred to pair them. This allowed us to iterate the taxonomy components (for instance, in distinguishing assignment delivery from submission) as well as the pairing rules (for instance, in distinguishing best-of the two from the average).

These considerations critically drive our ordinal, modular formulation of an LMS "Depth of Use" metric. Table 3.1 outlines the final rules for "low", "medium" and "high" use of seven Canvas services (announcements, syllabus, discussions, assignments, quizzes, gradebook, and files), and figure 3.3 describes how these rules are paired and aggregated to arrive at the overall DOU ranking for the course. The taxonomy in table 3.1 forms the basis of the course-level DOU measurement (in service of preliminary **Objective 3.A**), which is evaluated for its relationship with course participation, modality, logistics, and outcomes (in service of preliminary **Objective 3.B**). These two preliminary objectives come together to identify our primary research question for Study I (Section 1.3.1).

img/dou_estimation_diagram4.png

Figure 3.3: DOU Estimation: identifying DOU scores for individual LMS tools using the taxonomy identified in table 3.1, and assembling them (best-of/average) into the overall course DOU (equation 3.1).

Definition We define depth-of-use for an LMS resource R_i as a simple logic rule DOU_i of the form $(R == k_i)$ where k_i is a whole number. For instance, per table 3.1, $(\mathbf{An} == 1)$ for a given course implies *some* use of announcements (placeholders or class schedules, no instructor or TA activity). A total of N resource DOUs are accounted towards each course. As visualized in figure 3.3, the overall DOU for the course, DOU_c is *aggregated* from the resource DOUs as follows.

$$\mathrm{DOU}_C \triangleq \zeta(P_1, P_2, ..P_{M'}, S_1, S_2, .., S_{N'}) \tag{3.1}$$

where

$$P_{ij} = \beta_i \Big(MAX\Big(DOU_i, DOU_j\Big)\Big) + (1-\beta_i)\Big(\zeta\Big(DOU_i, DOU_j\Big)\Big) \tag{3.2}$$

img/dou_two_beta.png

Figure 3.4: A simple illustration of how two LMS resource DOUs are paired using equation 3.2. In this example, assignment delivery DOU is 1 (link to DOC, ZIP or app), and assignment submission DOU is 0 (no file upload, likely paper or app). (Left) Setting β to 0 computes the MAX of the two resource DOUs, which is useful when only one is needed for overall DOU. (Right) Setting β to 1 computes the floored average, which is useful when both need to contribute to DOU equally.

Equation 3.2 describes how two resource DOUs A and B are paired in P_i. We choose to apply MAX() or $\zeta()$ by setting β_i to 0 or 1, respectively. $\zeta()$ is the logic equivalent of a real-valued floored-average $AVG(X, Y)$ function. Figure 3.4 illustrates this step for assignment delivery ($\mathbf{A}_d == 1$) and assignment submission ($\mathbf{A}_s == 0$). $MAX()$ assigns the output to the larger of the two input contributions, hence $P_i^{(\mathbf{A}_d,\mathbf{A}_s)} = 1$. On the other hand, $\zeta()$ gravitates to the lower of the two, hence $P_i^{(\mathbf{A}_d,\mathbf{A}_s)} = 0$. Picking $\beta_i = 0$ implies that the instructional staff intends to consider the MAX(), or the *best* of assignment delivery and submission DOUs towards the overall LMS DOU. On the other hand, $\beta_i = 1$ rewards contributions from both DOUs equally, so *both* assignment delivery and submission need to utilize Canvas thoroughly for a high overall DOU rank. Finally, in equation 3.1, we average all of the pairwise (P), and single (S) terms using $\zeta()$ to compute a final score (low, medium or high) of overall LMS use. The findings in Section 4 are based on $\beta_i = 1$ for pairwise DOU terms ($\mathbf{A}_d$, $\mathbf{A}_s$) and ($\mathbf{Q}_d$, $\mathbf{Q}_s$), and $\beta_i = 0$ for pairwise DOU terms ($\mathbf{S}$, $\mathbf{F}$), and ($\mathbf{D}$, $\mathbf{G}$). Note that DOU allows flexibility in both *pairings* and *weights*, to encourage research on the usability and perceived efficacy of custom DOUs for a variety of tools and learning environments. In addition, table 3.1 is vendor-agnostic, in that it can measure the use of multiple LMS ecosystems, and taxonomies for LMS resources can be added or subtracted on a need-basis.

Three important practical considerations emerge in the design of DOU. First, courses frequently contain multiple delivery and submission types for assignments and quizzes. We thus require the overall DOU criteria in the taxonomy to hold true for a simple majority (at least 50%) of assignments or quizzes in the course. For instance, at least 50% of the course assignments should be fully hosted on Canvas for the assignment delivery DOU to assume a value of 2. Second, we define high discussions DOU ($\mathbf{A}_d$) with the presence of one or more live Canvas discussion threads (at least one post per week or course instrument), same as the announcements DOU (**An**). These heuristics aid the overall parsimony and interpretive power of the DOU taxonomy. Third, we do not incorporate issues of information quality and use-quality for specific LMS tools, such as relevance and degree of reflection in discussion posts, or the ease-of-use and diversity of assignment submission modalities (notebooks, error logs, images, hyperlinks). These aspects are important to evaluate in faculty's software use. However, they pose significant challenges related to data sparsity (wide variation in data availability across departments), need for domain knowledge (learning theories, objects, and environments in use), and lack of feature parity across LTI apps, making the design of a unified taxonomy extremely challenging. DOU is envisioned foremost as a platform-level metric of LMS use across departments, so we identify issues of information-quality and use-quality as outside the scope of the current taxonomy, and reserve them for future work.

Table 3.2: Key counts and DOU breakdown (% **Lo**, **Med**, **Hi**) for course cohorts in the spring 2017 dataset.

Attribute	#	%
Overall	6117	58, 29, 11
Undergraduate	4470	56, 30, 12
STEM	3730	60, 29, 9
Online	981	54, 31, 13
Third-Party App Use	2124	34, 47, 18
Digital Skills Training	2286	55, 33, 10
Viewership (μ, σ)	682, 5e4	-
Enrollment (μ, σ)	49, 96	-
#TAs (μ, σ)	0.4, 2.8	-

3.3 Datasets and Methods

The first two steps in this study relied on a series of brainstorming sessions with IT personnel (N=5), using aids visualized in figure 3.2. The notes collected during these sessions were synthesized by the research team (dissertation author and a graduate student collaborator) into design requirements using open coding and thematic analysis [107]. For the second step, we used the final two questions listed in Appendix E.1 to create an aggregate list of how all participants taxonomized the use of LMS services, and how they preferred to pair them. This allowed us to iterate the taxonomy based on the data received (for instance, in distinguishing assignment delivery from submission) and the pairing rules (for instance, in distinguishing best-of the two from the average).

The third step of this study was a quantitative analysis of course metadata collected for 50000+ courses during the fall and spring academic terms between the years of 2017 and 2022 from Canvas, the enterprise LMS in operation at Virginia Tech. It is important to note that a university-wide transition to online instruction in the spring 2020 academic term in response to the COVID-19 epidemic makes the 2020 dataset unreliable for longitudinal analysis, hence excluded from our study. Table 3.2 lists key aspects of the 6117 courses analyzed during the spring of 2017. For instance, 4470 (73%) of these courses are intended for undergraduate audiences, 3730 (61%) courses deal with STEM content, and 5136 (84%) use traditional, face-to-face instructional format. These majorities are also retained in each of the three DOU groups as per table 3.3, with important differences. We used a combination of manual and automated strategies (web scraping, entity resolution, and topic modeling) to create LMS utilization metadata for each course. Key textual sources include, and are not limited to, the Virginia Tech course catalog and historical timetable, Canvas page request logs, course descriptions on the Virginia Tech website [86], as well as syllabus files and assessment page content from Canvas course sites. STEM tagging of courses in the dataset is in accordance with the DHS classification of STEM fields [158].

To answer our study research question (**RQI**), we begin by testing our hypotheses (**H1** - **H10**). DOU is ordinal and not normally distributed, so we use non-parametric Kruskal-Wallis H-test [99], in addition to an independent two-sample t-test, for hypotheses with discrete-valued meta-variables. Table 3.4 and figure 3.5 describe the outcomes of these tests. We evaluate group differences in viewership and enrollment for each of low, medium and high DOUs using one-way ANOVA (F-test). To expand our analysis, we then claim-test each of the hypotheses against all constituent dimensions of DOU (tables 3.5 and 3.6). We combine these hypothesis tests with frequency and cohort analyses to examine the needs of the DOU use-case (adoption, support, learning outcomes).

Table 3.3: Spring 2017: High, medium and low DOU group composition (%) by course and instructor attributes.

DOU	**Undergrad**	**STEM**	**Online**	**App use**	**Skills**
Low	69	63	16	20	35
Medium	75	60	19	55	42
High	83	52	23	57	33

Table 3.4: Spring 2017: Hypothesis-testing the relationship between LMS DOU and key course attributes.

Hypothesis	t	F	H
H1: Undergraduate	8.3**	68.4**	63.9**
H2: STEM	-5.0*	25.1*	21.3*
H3: Online	3.6*	12.6*	12.1*
H4: App use	28.4*	8e2**	7e2**
H5: Enrollment	-	73.8**	614.4**
H6: Viewership	-	4.1*	0.9
H7: GPA	-	7.4*	9.0*
H8: DFW	-	6.9*	2.5
H9: #TA	-	97.9**	1e3**
H10: Skills	1.5	2.3	6.3*

*stat. signif., $\alpha = 0.05, p <= \alpha \wedge p > 1e{-}10, **\, p < 1e{-}10$

3.4 Findings

Modality (H1-H4)

As per Table 3.4, undergraduate courses have higher average DOUs relative to graduate courses (t-statistic is positive), consistent with their higher average enrollment (61 as opposed to 19 for graduate courses). As per table 3.5, undergraduate courses have higher relative DOUs for announcements ($F = 76.9, p < 0.01$), grading ($F = 119.8, p < 0.01$) and online syllabi ($F = 30.2, p < 0.01$), among others. Non-STEM courses feature higher use of the LMS for assignment delivery ($F = 32.8, p < 0.01$) and submission ($F = 24.1, p < 0.01$), among others. Traditional in-class instruction loses out to online-only courses in overall DOUs.

img/dou_avgs.png

Figure 3.5: Average DOUs by course attributes.

Online instruction is linked to in-depth use of online syllabi ($F = 12.5, p < 0.01$), as well as assignment delivery ($F = 14.9, p < 0.01$) and submission ($F = 14.2, p < 0.01$). Roughly 70% of each of low, medium and high DOU courses rely on third-party apps (table 3.3). Reliance on third-party apps coincides with the use of announcements ($F = 772, p < 0.01$), gradebook ($F = 864.9, p < 0.01$) and discussion forums ($F = 181.9, p < 0.01$), among others.

Participation (H5-H6)

Higher *DOU* courses feature larger overall enrollment ($F = 73.8, p < 0.01$) and viewership ($F = 4.1, p = 0.01$), as per table 3.4. Both of these are strong correlates of LMS utilization overall, and across a number of LMS resources considered individually (table 3.5). High enrollment is linked to high use of detailed online announcements ($F = 89.8, p < 0.01$), assignment delivery ($F = 8.06, p < 0.01$) and discussion forums ($F = 8.75, p < 0.01$), among others. High site viewership is similarly linked to the use of syllabi ($F = 3.2, p = 0.03$), assignment delivery ($F = 4.4, p = 0.01$) and gradebook ($F = 8.6, p < 0.01$), etc.

img/dou_regression_bound.png

Figure 3.6: Regression tests on DOUs scores with course attributes (modality, participation, logistics, outcomes) as independent variables.

Outcomes (H7-H8)

The average course GPA is significantly linked to overall DOU as per table 3.4 ($F = 7.4, p < 0.01$), and the use of announcements ($F = 5.5, p < 0.01$), syllabi ($F = 11.2, p < 0.01$) and discussion forums ($F = 6.6, p < 0.01$), among others. In comparison, DFW rate is a weaker correlate of DOU (compare the magnitudes of overall F- and H-statistics). Smaller DFW rates coincide with higher online quizz submission ($F = 15.1, p < 0.01$), announcements ($F = 4.7, p =< 0.01$) and gradebook ($F = 7.7, p < 0.01$), among others.

Logistics (H9-H10)

As per table 3.5, the number of teaching assistants is significantly linked to higher DOUs for announcements ($F = 89.1, p < 0.01$), discussion forums ($F = 12.5, p < 0.01$) and gradebook ($F = 179.8, p < 0.01$). Participation in an online digital skills training program run by Virginia Tech is not strongly linked to overall LMS use. It is nonetheless linked to higher resource DOUs for discussion groups, syllabi, and files.

Table 3.5: Hypothesis-testing: $|t|$ and $|F|$ magnitudes for the relationship between resource DOU and course attributes (announcements, syllabus, files and assignment delivery).

Hypothesis	**An**	**S**	**F**	$\mathbf{A}_d$
H1: Undergraduate	8.8*, 76.9*	5.5*, 30.2*	5.4*, 29*	-1.3, 1.8
H2: STEM	-1.5, 2.2	-5.8*, 33.4*	-1.1, 1.1	-5.7*, 32.8*
H3: Online	1.5, 2.3	3.5*, 12.5*	2.2*, 4.8*	3.9*, 14.9*
H4: App use	27.8*, 772*	29.3*, 855.7*	29.2*, 854*	0.4, 0.2
H5: Enrollment	- , 89.8*	- , 77.9*	- , 75.7*	- , 8.06*
H6: Viewership	- , 4.8	- , 3.2*	- , 7.4*	- , 4.4*
H7: GPA	- , 5.5*	- , 11.2*	- , 4.6*	- , 5.9*
H8: DFW	- , 4.7*	- , 11.5*	- , 5.1*	- , 4.2*
H9: #TAs	- , 89.1*	- , 79.4*	- , 117.1*	- , 0.74
H10: Skills	0.3, 0.1	3.8*, 14.6*	3.9*, 14.8*	-0.4, 0.2

*stat. significant, $\alpha = 0.05, p <= \alpha, F > F_{crit}$

3.5 Contributions and Implications for Practice

3.5.1 Helping Faculty Evaluate the Costs and Benefits of LMS Tools

Evident from literature surveyed in section 2, LMS adoption is a complex process, geared by the perceived quality of the overall system and the information it serves, as well as historical differences in pedagogies, and faculty-perceived opportunity-cost of transition [166]. While determining the relative contribution of each of these factors is an open research problem, evidence in section 4 puts the needs for scale, interoperability, and ubiquitous access among the most important potential correlates of LMS adoption.

Scale

As per figure 3.6, and hypothesis **H1** in table 3.5a, larger class size coincides with higher or 'deeper' use of announcements, most likely because mailing lists become increasingly

Table 3.6: Hypothesis-testing: $|t|$ and $|F|$ magnitudes for the relationship between resource DOU and course attributes (assignment submission, quiz delivery and submission, gradebook and discussions)

Hyp.	$\mathbf{A}_s$	$\mathbf{Q}_d$	$\mathbf{Q}_s$	G	D
H1	2.4*, 6*	1, 1	-0.2, 0.1	10.9*, 119.8*	-1.6, 2.4
H2	-4.9*, 24.1*	-0.7, 0.5	-1.8, 3.2	0, 0	-10.3*, 105.7*
H3	3.8*, 14.2*	1.2, 1.4	0.4, 0.2	2.9*, 8.3*	8.5*, 72.1*
H4	-0.3, 0.1	-0.8, 0.6	-0.8, 0.6	29.4*, 864.9*	13.5*, 181.9*
H5	- , 14.4*	- , 0.39	- , 4.82*	- , 160*	- , 8.75*
H6	- , 1.2	- , 0.02	- , 2.6	- , 8.6*	- , 0.72
H7	- , 2.2	- , 0.78	- , 15.1*	- , 7.7*	- , 6.6*
H8	- , 2.6	- , 0.3	- , 15.7*	- , 7.7*	- , 4.05*
H9	- , 2.6	- , 2.6	- , 0.5	- , 179*	- , 12.5*
H10	-2.7*, 7.5*	-0.6, 0.3	0.3, 0.1	-0.1, 0	2.6*, 6.7*

*stat. signif., $\alpha = 0.05, p <= \alpha \wedge p > 1e{-}10, **p < 1e{-}10$

inefficient and harder to organize and search at scale. Larger audience sizes also coincide with more frequent LMS use for assignment submission and delivery. One key reason is that this allows for a larger range of content to be submitted and greater flexibility in scheduling and organizing take-home exams and offline evaluations. In comparison, according to hypothesis **H4** in table 3.5, the use of third-party apps coincides with that of online discussion forums, but not for assignment delivery and submission. Services like Piazza are particularly favored by faculty because of their advanced forum management, content processing and tagging features, compared to the newer Discussions app aboard Canvas. Nonetheless, the largest effect sizes for **H5** (table 3.5) correspond to the use of gradebook, announcements, syllabi, and files. Undergraduate, non-STEM courses are likely to utilize these LMS tools. These courses are typically major-unrestricted, and enroll hundreds of students across multiple sections in a given academic term. Hypothesis **H9** (tables 3.5) suggests this also coincides with higher numbers of teaching assistants. Early adopters in the instructional staff of these courses especially gravitate towards basic housekeeping use-cases for LMS tools, such as communicating class times, office hours, course milestones, and grades, whilst retaining their use of third-party apps (**H4**). Faculty's ability to delegate administrative and technology discovery tasks can thus critically help them balance their research and teaching duties and potentially migrate to new tools as class sizes increase. Third-party apps with free tiers, local authorship, and open-source communities remain consistently popular because of their low overhead of initial setup. But, without adequate access to teaching assistance, scaling the use of these apps to high-enrollment classes, managing student feedback, and providing timely technical support are likely to remain challenging.

Interoperability and Ubiquitous Access

Intuitive, safe, and swift data transfer between educational apps is essential to minimizing faculty's cognitive burden-of-discovery and strengthening institution-wide LMS adoption rates. For instance, the enduring utility of Canvas's file and assignment/quiz management apps observed for Virginia Tech faculty is in part because of their easy integration with grading apps. This lets course staff grade assessments without worrying about manual data imports or data corruption. Figure 3.7 describes the frequently-used third-party apps at Virginia Tech, Department of Computer Science, and the specialization areas of corresponding courses. The commonly used services in these app-suites are discussion forums and course content management (Piazza, Top Hat), exam management (WebCAT), programming instruction and interactive visualizations (OpenDSA, BlockPy, CodeWorkout), etc. Used frequently often by undergraduate courses on programming, algorithms and software engineering, these apps do not affect course GPA and DFW rates (considered together or individually) in the department. While they offer seamless integration with LMS tools for course, student and exam management, many of these apps lack one-to-many LTI connections which allow cross-course access, collaboration and research features. This limits the adoption of these apps beyond their parent departments and research groups [143]. Lack of img/app_shares_with_legend.png

Figure 3.7: Third-party app use at Virginia Tech by frequency of use and computing specialization areas. Discussion forums and course content management (Piazza), and programming instruction (OpenDSA) combined account for nearly half of all third-party app use in Spring 2017.

essential interoperability and ubiquitous access features in an LMS (such as reusing legacy materials in future iterations of the course) often restricts its use to housekeeping functions (modest or one-off use of announcements, files, and gradebook), discourages research into new pedagogies, and fuels poor returns on the institutional capital investment into LMS tools.

3.5.2 Facilitating Institutional Support and Management

LMS administrators and instructional designers can use DOU to support departmental resource allocation, faculty development, course evaluation, and LMS evangelism. In this section, we discuss our experiences with these DOU use-cases at Virginia Tech, and the lessons learned.

Table 3.7: Low-DOU course frequencies by context

Context	%	Context	%
Lo #TA ∧ No app use	79	Grad ∧ Online ∧ Lo #TA ∧ No skills	71
Lo enroll ∧ Lo #TA ∧ No app use	70	Grad ∧ Lo #TA ∧ No skills	67
Undergrad ∧ Lo #TA	67	Grad ∧ Online ∧ Lo #TA	65

Resource allocation and professional development

DOU can inform the allocation of teaching support at the department or college level. It can serve as a data-driven signal of the need for direct, personalized interventions or additional teaching support for faculty micro-cohorts. For instance, in table 3.4, the hypothesis **H10** brings the relative utility of a comprehensive professional skills program into question (compared, for instance, to number of TAs in **H9**), as the cohort is at best indifferent to 'deeper' LMS use. A similar picture emerges in table 3.7 where high enrollment courses with little to no teacher support staff results in a substantial fraction of low DOU courses (79.3%). The availability of digital skills training does not affect the wide majority (about 70%) of these courses. Low DOU courses often frequent the cohorts with low #TAs, and faculty training alone does not appear effective in alleviating the cognitive burden of discovery required for rapid adoption. We also observe, for instance, that according to hypothesis **H7** (table 3.5), higher average course GPAs and lower DFW rates are linked to higher quiz submission and syllabus DOUs. Undergraduate management, leadership and policy courses make up 35% of the course cohort with no online quiz submissions, while STEM undergraduate courses make up a majority of high Canvas quiz submissions. Natural resource management (graduate) and physics (undergraduate) courses make up 40% of courses with no online syllabi. Identifying micro-cohorts with deficiencies in app-level LMS use can help colleges and departments develop online training tools, allocate technical support esp. during a pandemic, and improve outreach and faculty buy-in towards new LMS tools.

img/courseab_new3.png

Figure 3.8: LMS page request volume and types for the 'junk-drive' use-case. Page requests for course A (top) occur around three key course milestone deadlines, and reflect a variety of LMS resources (backup quiz submissions, show courses, create quizzes, show course activity summary, show quizzes). Course B (bottom) uses the LMS site as a storage drive with no student activity, with nearly all page requests belonging to a single category (show files).

Design assessment and LMS evangelism

DOU can point LMS administrators to faculty preferences about the use of LMS tools and legacy apps, and their broader reasons like trade-off between teaching and research responsibilities, faculty self-efficacy, and cognitive burden of discovery [31, 166]. It can, therefore, inform the development of effective, closed-loop processes for (1) course redesigns, (2) technical support allocation, and (3) software evangelism on campus.

Table 3.7 describes some example low-adoption cohorts which highlight the connection be-

Table 3.8: Compositions of the junk-drive, gradebook-only and access-portal course cohorts in Spring 2017. For instance, 42% of the gradebook-only cohort has a large class size, and 80% has low #TAs (if any).

Course attribute	Junk-drive (%)	Gradebook-only (%)	Access portal (%)
Undergraduate	63	77	66
STEM	65	66	60
Online	12	12	15
3rd-party app use	18	58	100
Enrollment (Lo, Hi)	78, 21	57, 42	84, 16
Viewership (Lo, Hi)	36, 63	96, 3	98, 1
#TAs (Lo, Hi)	97, 2	80, 19	95, 4
Skills training	56	40	39

tween student viewership, course modality and logistics (low DOU courses are 58% of the dataset overall). In an expert review session with five instructional designers at Virginia Tech, we identified four distinct types of low-adoption use-contexts or "course personas", and their implications for design interventions: *junk-drive*, *gradebook-only*, *access-portal*, and *housekeeping*. We review them as follows:

Junk-drive According to table 3.2, the overall frequency of low-DOU courses in the dataset is 58%. Compare these with the frequency of low-adoption courses for several micro-cohorts in table 3.7. These frequencies echo the connection between instructor and student engagement and how key aspects of course content and logistics might affect the system and information quality experienced by students while interacting with the LMS.

An interesting scenario emerges in the connection between viewership and DOU. Figure 3.8 visualizes the weekly average pageviews for two STEM courses with medium weekly viewership and vastly different DOUs. The share of page requests by category (application controller::action) reveals the differences in LMS utilization: course B is primarily being used as a file drive despite having gone through design intervention. Course A, on the other hand, reveals heavy LMS use around two key deadlines for the course and a surge in page views early on in the semester (corresponding to add-drop period for the term). We identified 114 low-DOU high-viewership courses in the spring 2017 dataset. Table 3.8 details their attributes. They are slightly more likely to be undergraduate and STEM courses, and despite about half of them reporting digital skills training, only about 18% report the use of third-party apps. Nearly all of these courses do not have teaching assistance and the class sizes are mostly small, so there is evidence of instructor use or experimentation, however preliminary, with native LMS services. This micro-cohort is an important example of the potential for continual LMS evangelism and instructional support in order to drive up adoption rates.

Gradebook-only According to table 3.8, exclusive use of the Canvas gradebook likely coincides with medium to high-enrollment, undergraduate (77%), STEM (66%) courses, with heavy reliance on third-party apps (58%), and an abundance of labs, recitations and group projects. Digital skills training is particularly ineffective for this cohort, which brings to attention its scarce teaching support staff (80%). It simultaneously points to the need for design interventions that help reduce the cognitive burden of faculty looking to make a fuller transition to LMS discussion forums, groups and assessments, especially at scale.

Access-portal This micro-cohort refers to course sites that are collections of links to third-party apps. Per table 3.8, these courses are often undergraduate, STEM and unresponsive to digital skills training. Such an extreme reliance on these apps is often a function of both department-level precedents and faculty-perceived ease-of-use. This implies that a design intervention for this micro-cohort should make a particular note of faculty's technology self-efficacy and access to teacher support (note the high fraction of low #TA courses) in end-of-semester quality assessments.

Table 3.9: COVID-19 pandemic response planning: overall course DOUs before and after the institution-wide transition to emergency remote teaching at Virginia Tech (Spring 2020).

DOU	**Before transition to remote teaching**	**After transition to remote teaching**	**% Change**
Low	3932	3667	-6.49%
Medium	2365	2288	-3.26%
High	673	1004	+49.18%

Housekeeping This cohort of Canvas courses features heavy use of announcements, files, and gradebook, with third-party apps for other course functions. These courses represent a considerable fraction (nearly half) of Canvas late-adopters during the COVID-19 emergency remote teaching era, and likely point to a lack of technical support or faculty's low bandwidth for outreach, especially beyond the IT/TLOS self-help resources.

Table 3.10: COVID-19 pandemic response planning: % change in course DOUs by LMS tool after institution-wide transition to emergency remote teaching at Virginia Tech (Spring 2020).

DOU	**An**	**D**	**F**	**S**	$\mathbf{A}_D$	$\mathbf{A}_S$	$\mathbf{Q}_D$	$\mathbf{Q}_S$	**G**
Low	-7%	-9%	-4%	-1%	-13%	-54%	+20%	-28%	-4%
Medium	-23%	+63%	-	-0.7%	-4%	+12%	+27%	-	-
High	+44%	+40%	+6%	+2%	+15%	+24%	+30%	+36%	6%

3.5.3 Planning the Pandemic-Era Transition to Remote Teaching

In the spring of 2020, an institution-wide policy of emergency remote teaching was rapidly enacted by Virginia Tech IT leadership, in response to the COVID-19 pandemic. System administrators began with a DOU analysis conducted at the beginning of the term to determine key low-DOU course clusters (upper level STEM and general education coursework), frequent high DOU LMS features (typically the ones with lowest cognitive burden-of-discovery like files and gradebook), and frequent low DOU LMS features (quiz and assignment delivery).

The administrators facilitated a rapid transition to remote teaching over a period of two weeks by focusing their support on low-DOU instructors. They designed training sessions, in-person consultations, and in-depth documentation focusing on delivery and submission of assignments and quizzes via Canvas. According to table 3.9, the IT transition team was able to increase the total number of high DOU courses by over 49%. Table 3.10 breaks down the post-COVID DOU gains by LMS tools. Three key takeaways emerge. First, the smallest gains were observed in files and gradebook modules, which hints at an abundance of junk-drive and gradebook-only courses prior to the transition. Instructors new to an LMS tend to first explore the tools they can utilize without significant cognitive effort. Second, the team found that post-transition, instructors' use of announcements (for bulk communication within the LMS) and discussion forums (as a replacement for in-class, face-to-face interaction) increased significantly (+44%, and +40%, respectively). Third, the increase in courses with high DOUs for assignment delivery (+15%), assignment submission (+24%), quiz delivery (+30%), and quiz submission (+36%) is often at the expense of low DOU courses in the same category. This suggests that in favoring online course assessments, Virginia Tech faculty responded to the transition team's focused development and support initiative. While an uptick in overall DOUs is expected during a transition to remote teaching, we contend that if the transition team had not been able to marshall their resources and provide directed support based on DOU analysis, we would observe an abundance of low → medium DOU growth. The instances of low → high and medium → high DOU growth suggest that our pre-transition analyses facilitated an actionable assessment of key faculty needs and a focused adjustment of the IT training and support regimen which maximized its impact.

To summarize, table 3.11 notes the potential areas of impact DOU-enabled artifacts we present in this chapter for key impact areas of institutional support: instructional design, faculty outreach, and support allocation. The UDC dashboard in table 3.11 refers to a

Table 3.11: DOU impact areas for institutional support.

Impact Area, Job Tasks	**Design and Research Objectives**	**DOU-Enabled Artifacts+Analyses**
Course redesign, retooling, program certification	Developing or evaluating a course, course website, educational technology, or learning environment	UDC dashboard department/course personas causal inference decision matrices
Software rollouts, tech evangelism	Developing or evaluating faculty outreach and communication strategies	UDC dashboard structured rollouts personalized campaigns, conversion analytics, IT CRM, callback prediction
Support allocation, project management	Designing the balance of institutional investment in support vs. automation: procurement, staffing	UDC dashboard department/course personas, multi-view ROIs

`img/dou_dashboard.png`

Figure 3.9: A development instance of the UDC Canvas Depth of Use dashboard, with graphical (left) and tabular (right) views.

dashboard summarizing the department-level and course-level DOU metrics, and their implications for department leadership currently in development at Virginia Tech (TLOS), with a production target of early Spring 2024 (figure 3.9).

3.6 Limitations and Future Work

Study I contends that a multitude of software services, and large volumes of user data make it difficult to assess and improve the usability of LMS-hosted educational apps, support pedagogy, and design meaningful interventions for at-risk students. To solve this problem, a critical first step is to concisely describe how these software services native to an LMS are utilized in aggregate by a university course. Study I thus proposes a novel method (**DOU**) to convert expert-sourced taxonomies of LMS use into a single ordinal (low, medium or high) ranking. The study then evaluates the relationship of LMS use and course attributes such as modality, participation, logistics, and outcomes for nearly 20,000 courses between 2017 and 2019 to discover frequent use-contexts (study methods detailed in our ITiCSE'20 paper).

The hypothesis tests in Study I demonstrate that the **DOU** framework enables evidence-based discovery of stakeholder needs in connection with the use of technology. For instance, we discover that needs for scale, ubiquitous access and interoperability drive faculty use of LMS tools. We also discover that DOU is frequently linked to better aggregate course outcomes (GPA, DFW), teaching support, and third-party app use. Finally, we conclude that instructional designers, LMS administrators and department leadership can use DOU-based analyses to inform LMS evangelism, design interventions, professional development, and pandemic response planning.

In depth-of-use, we devise a multi-factor, resource-specific view of LMS utilization. DOU helps us examine a variety of use-contexts in faculty and student adoption of LMS services.

Our hypothesis-testing reveals that the needs for scale, ubiquitous access and interoperability drive a broad swath of courses across departments towards higher LMS use. We also discover that DOU helps us isolate low-adoption course cohorts, allocate institutional support, and reflect on faculty preferences, technology limitations, and administrative policies that might drive these cohorts. Our research aims to combine expertise from course planning, policy design and quality assurance in order to test multi-level claims of efficacy and recommend interventions that leverage the totality of contextual evidence of historical LMS use.

Our dataset and analyses describes all Canvas course sites commissioned during the fall and spring terms of 2017 through 2022 at Virginia Tech. Its scope can be broadened in several important ways. We examine these as directions of future work as follows. To aid generalizability, we intend to reproduce our analyses for Scholar LMS - in use prior to Canvas - at Virginia Tech. We also plan to compare our results with courses hosted aboard Canvas at peer institutions. We hope to hypothesis-test DOU as a function of course modality (flipped and blended classrooms [40]), and content and system quality (example pervasiveness [163], cognitive task models [111, 134], early availability of course content, site aesthetics [110], mobile platform support [26, 27] and accessibility [167]), specialized learning environments, and funding inequities at the department and college level in order to analyze their impact on the usability of LMS services.

We intend to expand the characterization of LMS use by resource (table 3.1) to include the use of content recommenders. The domain of educational recommendation has a large volume of literature on highly specialized interventions aimed at a multitude of use-contexts [104]. Incorporating the use of topic, course and supplementary content recommenders in DOU can help evaluate if a specific low-adoption cohort is responsive, in perceived ease-of-use, novelty, trust and satisfaction [73, 75], to instructor-aided curation of study materials. We also plan to account for user-activity within third-party apps hosted by the LMS. We plan to collaborate with several app vendors to better understand the relative satisfaction with interactional and content quality these apps might provide.

In recent years, educators and IT administrators have been widely interested in the use of emerging tools like generative AI [105], virtual and mixed reality [135], and short-form video [48] to support higher learning. These technologies present a promising array of use-cases in teaching and learning, such as supporting sensemaking [153], productivity [161], groupwork [85], and assessment [154]. We identify the development of DOU taxonomies for these technologies, and validating their relationship with learning outcomes, as crucial vectors of future work. These taxonomies can help instructors evaluate teaching efficacy, and provide IT administrators with decision information on pilot testing, budgeting, licensing, and infrastructure management for new software. We also envision a broader role for LMS-hosted content recommender systems [104] as vehicles for faculty outreach, micro-learning, professional development, and personalized technical support. We seek to evolve the DOU measurement and validation strategies in this study to support these emerging technologies at scale.

Finally, the scope of our analysis is interpretive in that it examines the observed LMS usage as a function of high-level course meta-characteristics. In our future work, we plan to concurrently model instructor preferences, habits, and values that make up the said usage. We plan to incorporate instructor work experience and familiarity with instructional design practices in our analyses. We also plan to collect feedback from instructors and students, using semi-structured interviews and online academic forum analyses [76] for key low-adoption micro-cohorts to better summarize and validate these reasons.

3.7 Chapter Summary

In this chapter, we propose "Depth of Use" (DOU), a metric which summarizes course faculty's use of the learning management system (LMS) for instructional support providers. It can help identify the needs and constraints driving faculty's use of educational technologies. It can also help manage the institutional support resources available to meet these needs. Finally, it can inform the design of sensemaking and decision-analytic artifacts to support the work of a multitude of higher education stakeholders, such as department leadership, IT administrators, and project managers.

Chapter 4

Understanding Faculty's Editorial Intent and Trust Perceptions for Educational Recommendation

Generating faculty buy-in in a novel recommender system (RS) for technology-enhanced learning and instructional support is a non-trivial challenge. Previous literature on adoption of information systems has long reported that humans treat recommendation agents as "social actors" and trusted objects [15][170], and this trust is inextricably linked to system ease of use, degree of control, algorithm transparency, and trust in automation [136][84]. User experiments with RSs typically assess the *experiential* notion of trust using the *perceived* trustworthiness of the RS output. Algorithms for recommendation also frequently incorporate explicit and implicit signals of trust, for instance, the RS user's *interactional* awareness of their local neighborhood, or some consensus of the preferences of their self-reported, trustworthy friends in a social network at large [113]. While algorithmic awareness of a user's neighborhood is important for producing accurate recommendations, real life recommendation domains often involve user groups with differences in institutional, group, or task-based roles, powers and prerogatives. Domains like higher education, sacred spaces, and news allocate editorial prerogatives to a subset of stakeholders considered to be owners, gatekeepers, or arbiters of domain knowledge [97][78]. For instance, faculty are considered responsible for executing the content policy for their course (textbooks, supplementary readings, audio/video, software use, data practices) while observing the guidelines set by department leadership, and can invite participation from teaching assistants and students in these decisions [78]. IT administrators recommend, but do not typically compel the adoption of educational software on faculty [42].

We therefore argue that recognizing the editorial power relationships between stakeholders in the RS application domain is one way to begin to identify the broader context of trust in the recommendation algorithm, and to expand the interpretive power of the RS output. Study II examines Virginia Tech faculty, teaching assistants, and students' preferences of editorial authority and trust in algorithms, for a hypothetical 'Suggested Readings' recommender system aboard a learning management system (LMS) course site. Using a simple, first-principles metric of editorial authority (**E-Auth**), we hypothesis-test the relationship between RS editorial task distribution and stakeholder trust. We then describe the top three editorial roles (author, active viewer, viewer) allocated to students by course staff, and identify their frequent contexts, rationales, and RS use cases. Figure 4.1 provides an overview of our study methodology, and the group attitudes and trust relationships we investigate.

4.1 Study Objectives

Outlined in the primary research question of Study II (section 1.3.2), the objective of Study II is to understand the editorial preferences (trust intentions) of faculty, teaching assistants, and students as a function of trust, specifically in the context of use of an educational recommender system. We aggregate four key RS tasks (seed, edit, refresh, and delete) using the notion of *editorial authority* and measure its allocation among faculty, staff and students. We then use analysis of variance (ANOVA) hypothesis tests to understand the relationships img/studytwo_overview.png

Figure 4.1: Study II overview: Recommendation stakeholders and their mutual trust beliefs (bottodm), editorial trust intentions (top-left) and study methods (top-right).

between editorial authority and stakeholder trust. We contend that these differences in *editorial authority* - exemplified in an editor-consumer relationship between faculty and students, for instance - are related to the trust both assign to each other and the RS algorithm. For instance, faculty's willingness to incorporate student and TA feedback into the RS algorithm can point to a belief in editorial authority for multiple stakeholders, or regard for automation in the longer-term.

4.2 Study Design: Editorial Authority, Trust and Algorithmic Agency

4.2.1 Definitions

We define the 'Editorial Authority' (**E-Auth**) allocated to a study participant as a linear aggregate of all editorial powers (seed, veto, rate, comment) they identify for their own user group relative to all the other user groups. The individual editorial powers are represented by binary ('yes' or 'no') votes. Consider equation 1 as follows.

$$\begin{gathered}\textbf{E-Auth (Faculty} \rightarrow \textbf{Students)} = \textbf{Seed Score} + \\ \textbf{Veto Score} + \textbf{Rate Score} + \textbf{Comment Score}\end{gathered} \tag{4.1}$$

For example, a faculty member's choice to let the TAs and students seed RS articles will result in a **seed score** of 2 (as in, the power to source recommended articles is shared with two user groups). If this faculty member favors exclusive veto power for faculty, but retains the rate/comment power for TAs and students, the **veto score**, **rate score** and **comment score** would be 0, 2 and 2 respectively, with a final **E-Auth** score of 6 (normalized to a percentage

Table 4.1: Key attributes of Study II survey participants

Role	#	# Male/Female	# Departments	# STEM
Faculty	27	13/14	16	17
Teaching assistant	6	4/2	5	6
Student	9	7/2	4	7

for our analyses). A small **E-Auth** score thus indicates that the course instructor might minimize student participation to reduce the cognitive effort of managing student feedback at scale, to prevent spam and inappropriate content, and to better align the course with learning outcomes in a given degree specialization. In Study II, we evaluate the relationship of **E-Auth** with participants' trust attributions (single-item) towards course staff, students, the recommendation algorithm, and automation.

4.2.2 Datasets and Methods

Study II was conducted in two steps (figure 4.1, study methods). The first step is a survey with responses from 42 participants (27 faculty, 6 teaching assistants, 9 students). We used single-item trust belief questions in this survey (5-point response item ranging from low to high, normalized to a percentage). Survey questions are noted in Appendix D.2. The second step is follow-up semi-structured interviews with 11 (6 faculty, 3 teaching assistants, 2 students) of the aforementioned 42, all affiliated with Virginia Tech, and interview questions are listed in Appendix E.2. Table 5.2 details the attributes of Study II survey participants. These faculty members, students and teaching assistants represent 16, 5 and 4 departments, respectively. 30 of the survey respondents (17 faculty members, 6 TAs, 7 students) represent STEM disciplines of study. We recruited participants on a rolling basis between August 2020 and July 2021, using convenience sampling and voluntary response sampling on departmental mailing lists and Facebook groups.

img/seed_veto.png

Figure 4.2: Editorial authority: Virginia Tech faculty, TA and students' opinions on who should be allowed to seed (left) and veto (right) articles for a 'Suggested Readings' RS. 44% of surveyed faculty allocated the article sourcing task to students, but only 4% preferred students participate in article vetoing.

For our analyses, Study II proposes a mixed-methods approach. We conducted hypothesis tests regarding the relationships between our study variables (relative trust in algorithms, automation, editorial authority) using one-way ANOVA (F-test). We also performed content analysis [107] on survey responses and interview transcripts to identify frequent themes in the study participants' commentary regarding their preferred RS editorial tasks for each user role.

4.3 Findings

Figures 4.2 and 4.3 illustrate key distinctions in the way Virginia Tech faculty, TAs and students allocation RS editorial tasks to each other in the Study II survey. In figure 4.2, nearly all faculty members allocate sourcing and vetoing tasks to themselves, but fewer (52%) think TAs should seed RS articles and even fewer (44%) allocate this task to students. The divide is quite apparent for vetoing, as all but one (4%) faculty member suggest that students should not be able to instantly remove recommended readings. According to figure 4.3, about 75% of faculty (as opposed to all students) favor the idea of soft power for students: the img/rate_comment.png

Figure 4.3: Editorial authority: Virginia Tech faculty, TA, and students' opinions on whether students should participate in rating and commenting on the output of a 'Suggested Readings' RS. For instance, 25% of surveyed faculty do not favor that students' ratings influence recommended articles, and 13% do not favor student comments.

img/trust_allocs_studyii.png

Figure 4.4: Average trust of course staff, students, RS algorithm and automation. For instance, an average student trust of 50%, 33%, and 52% is expressed by faculty, TAs, and students, in that order.

ability to rate individual recommendations and let the algorithm update the relative location of a recommended reading using group consensus strategies. 13% of faculty members and 17% of teaching assistants suggest that they do not favor students' ability to submit feedback about a recommended reading. According to figure 4.4 which illustrates trust relationships between stakeholders, all survey participants allocate more trust to course staff than they do to students and the RS algorithm, and faculty's trust of automation is slightly lower than that of students.

Table 4.2: Key results from hypothesis tests about the relationship between **E-Auth** scores (RS editorial authority assigned by faculty to students) and staff/student attitudes. 'Course staff' refers to faculty and teaching assistants considered together. For instance, per **H1**, teachers trust the RS algorithm less than teaching assistants, but according to **H2**, course staff does not trust the RS algorithm any less than students do.

Question Category	Hypothesis	$\|F\|$
Trust in RS algorithm	**H1**: Teachers trust the RS algorithm less than TAs.	6.49*
	H2: Course staff trusts the RS algorithm less than students.	0.19
	H3: All participants trust the staff more short-term.	10.8**
	H4: All participants trust the RS algorithm more long-term.	6.3*
Automation	**H5**: Staff trusts automation less in the short-term than students.	9.7**
	H6: All respondents trust automation more long-term.	3.9
	H7: Course staff trusts automation more long-term.	7.2**
Editorial authority	**H8**: Higher student trust long-term is linked to higher **E-Auth**.	3.7*
	H9: Lower average algorithm trust is linked to higher **E-Auth**.	6.5**
	H10: Favoring high automation is linked to higher **E-Auth**.	0.52

*stat. signif., $\alpha = 0.05, p <= \alpha \wedge p > 1e{-}10, **p < 1e{-}10$

Table 5.3 describes the results of hypothesis tests on key claims about trust in RS algorithm, preference for automation, and aggregate editorial authority (see research questions **RQ1-3** in section 3.2). We discover that overall, course staff (faculty and teaching assistants) tends to be more risk-averse in its trust relationship with RS algorithm and automation than students, favoring substantial human intervention in the sourcing, updating and removal of recommended readings in the short-term. Here is a detailed breakdown of these results.

Trust in RS algorithm (RQ1, H1-H4)

We discover that teachers tend to trust the RS algorithm less than teaching assistants ($F = 6.49, p = 0.01$). All user groups favor the role of course staff in ensuring trustworthy recommendations short-term more than long-term ($F = 10.8, p < 0.01$). Similarly, all user groups trust the RS algorithm more long-term compared to short-term ($F = 6.3, p = 0.01$).

Table 4.3: RS editorial roles allocated to students. Editorial tasks essential to each role are in bold text. Allocation refers to the number and fraction (%) of participants who chose to allocate a given editorial role to students. Average **E-Auth** refers to the amount of editorial authority assigned to students in each role. For instance, 39% of course staff favored students provide input as active viewers (AV), assigning them 54% of the available editorial authority.

Editorial Role	Editorial Powers/Tasks	Average **E-Auth**	Overall Allocation	Staff Allocation
Editor (**E**)	**Seed**, **veto**, rate, comment	80%	1 (2%)	1 (3%)
Author (**A**)	**Seed**, rate, comment	77%	17 (40%)	12 (36%)
Active Viewer (**AV**)	**Rate**, comment	54%	17 (40%)	13 (39%)
Viewer (**V**)	Comment	14%	7 (16%)	7 (21%)

Automation (RQ2, H5-H7)

Course staff tends to favor automation in recommendations less relative to students, especially in the short-term ($F = 9.7, p < 0.01$). Similarly, course staff favors automation in the long-term more than in the short-term ($F = 7.2, p < 0.01$).

Editorial authority (RQ3, H8-H10)

Higher trust in students is linked to preference for higher student **E-Auth** scores in the long term ($F = 3.7, p = 0.03$), hinting at a connection between perceived trust and editorial role assignment. Similarly, lower average trust in RS algorithm is linked to preference for higher student **E-Auth** scores ($F = 6.5, p = 0.01$). Preference for higher RS automation overall is however, not linked to editorial authority ($F = 0.52, p = 0.67$).

4.4 Contributions and Implications for Practice

4.4.1 Identifying Faculty-Preferred Editorial Modes for Students

In this section, we describe the top three frequently recommended editorial roles (**author**, **active viewer**, **viewer**) for students, as allocated by our study respondents (table 4.3) and examine their stakeholder trust (figure 4.5). Allocation refers to the fraction (%) of participants who chose to allocate a given editorial role (and corresponding powers and tasks) to students. For each editorial role, we also discuss the frequent themes (in bold text) of the mediating trust relationships we investigate in **RQs I** through **III**. These emerge from our content analysis, and they include, and are not limited to, differences in perceived expertise, supporting learner engagement, and scaling the learning environment. These editorial roles highlight distinct - if overlapping - user preferences that can suggest RS use-cases for supporting both teaching and research.

"The author" (A): Everything except veto

According to table 4.3, 40% of participants overall, and 36% of course staff favored the 'author' (**A**) role for students. **A**-faculty members prefer sharing article sourcing, rating and feedback responsibilities with students, but retain the veto power (i.e. the ability to instantly remove a recommended reading for all RS users) for the primary course instructor, as well as for the teaching assistants in a subset of cases. Survey respondents who prefer this editorial role also trust automation more than other roles (figure 4.5). Table 4.4 lists the frequent themes and course contexts for all RS editorial roles. In addition to citing their prerogative being in charge of facilitating student learning, **A**-faculty members frequently cited the **need to moderate content** and remove readings deemed irrelevant or malicious. According to img/eauthxtrust.png

Figure 4.5: Average trust as a function of preferred editorial roles allocated to students (editor, author, active viewer, and viewer). For instance, staff trust is higher than algorithm and student trust across editorial roles; and RS 'viewers' express an average student trust of 30%, whereas RS 'editors' express an average student trust of 61%.

one instructor: "*I worry about abusive behavior, unless it was clearly tracked*" and this would "*ensure the integrity of the course and the suggested readings*". They simultaneously acknowledged the potential utility of **student engagement** in these editorial tasks, and inquired about the ability of the RS algorithm to comprehend student needs early on (**H1**, **H4**), especially in graduate, research-focused courses. One faculty member said that based on her teaching experience, she perceived MS-level students to be resourceful and another commented:

> "*.. if objectionable material was posted, I'd rather we use it as a discussion topic rather than outright reject it.*"

Soft power, as in the ability to rate (or 'like' and 'dislike') content as a signal to RS algorithm, was frequently favored for students as a means of further engaging them with the course content. An instructor said of soft power, *"I like the way the soft power idea will engage*

Table 4.4: Frequent course contexts, themes, and use-cases for RS editorial roles allocated to students. For instance, study participants favored cooperative RS editorial models (**E/A**) for small class sizes (typically graduate courses). Courses with a preference for lesser or no student input (**AV/V**) have large class sizes (often undergraduate) and time constraints defined by course content (STEM, professional diplomas).

Editorial Role	Frequent Course Context	Frequent Themes	RS Use Cases
E/A	Graduate, Non-STEM	student engagement, content moderation	teaching
AV	Undergraduate, STEM	student disengagement, scale, outcome bias, content moderation	teaching, professional development
V	Undergraduate, STEM	student disengagement, scale, content moderation	research, teaching

their minds and (aid their) learning". This is consistent with hypothesis **H8** and **H9** in table 5.3. Teaching assistants often brought up the case of large course sections with multiple instructors, and how TAs and students having article sourcing powers could potentially help scale the teaching resources faster, and make them accessible to a larger subset of the student population. According to one teaching assistant,

> ".. *course instructors may sometimes not know when to remove content unless they are intimately familiar with it and know it to not be useful.*"

Student participants in our study were, by-and-large, in agreement with course staff about exclusive veto power for faculty and teaching assistants, especially because they felt that the ability to seed, rate and comment on articles gave them ample opportunity to engage with the course materials and RS authorship policies. As per one student: *"I feel anything that is against the teaching of the class should be allowed and a respectful discussion should occur; preferably in a forum setting."*

"The active viewer" (AV): Rate and comment

Table 4.3 suggests that 40% of study participants overall, and 39% of course staff favor an 'active viewer' role for students. This allocates the RS rating and commenting tasks to students while reserving the sourcing and vetoing of suggested readings for course staff. Members of this group frequently talk about the challenges of managing and responding to feedback at **scale**, especially in undergraduate courses with multiple sections and hundreds of students. One faculty member remarked about students not being able to seed or veto suggested readings:

> "*.. (this is) just to be able to manage with a course that has 14-15 sections of 65 students each.*"

This comment is not a lone anomaly. Over the last two years, the average undergraduate course at Virginia Tech has 3.7X the class size of the average graduate course (66 and 17 students on average, respectively), with many first-year, major-unrestricted courses enrolling several hundred students in any given term. It is worthwhile noting that **AV**-faculty members do not seek a fully cooperative model of student feedback to solve this challenge in the manner of **A**-faculty. Frequent reasons posited by faculty include perceived historical patterns of **student disengagement** and exclusive attention by students to course outcomes, or the **outcome bias** [73]. This bias is also known to affect student perceptions of instructors on academic social forums like *Koofers* [1] and *RateMyProfessor* [2]. As one faculty member commented,

> "*My experience has been that most students do not do more than what's required of them, unfortunately. They just want to know what they can do to earn an A in class.*"

Another faculty member talked about the requirements of professional diplomas - dictated by the rapidly evolving demands of the job market - as one key driver of student disengagement. He noted that the "*fast pace and skill-focused curriculum*" of professional diplomas (as opposed to research-based degrees) made it difficult for students to spend time on optional course content. In his experience, this had led to a substantial decline in student interaction with the course LMS site, to the point where cooperative editing of recommended readings seemed ineffective. In comparison, teaching assistants and students who favored the **AV** role largely cited course staff's **need for moderating** the RS articles for malicious behavior and honor code violations. About the article rating task, one student commented that it will "*allow the instructor and TAs to see what's most widely accepted*". Consistent with hypothesis **H7** (table 5.3), all participants expected the role of automation to be significantly higher long-term in consolidating student updates to the recommendations' rank order.

"The viewer" (V): Comment only

Study participants in this group favored the least egalitarian editorial model for recommending readings to students, and they have the lowest trust of students and automation relative to other groups (figure 4.5) . 16% of participants overall, and 21% of faculty and teaching assistants prefer no direct input from students in deciding the source and rank order of recommended readings. Same as **AV**, this is frequently observed for faculty members with undergraduate teaching responsibilities. This model considers student participation in article sourcing, updates and removal to be unsustainable if not counter-productive. Faculty survey respondents and interviewees frequently cited challenges of **content moderation** at **scale**. One instructor of an undergraduate Computer Science class described the problem of cooperative RS editing as analogous to the challenge of regulating discussion forums posts for the 200+ enrolled students in her section. Drawing on instances of age-sensitive commentary by students about an automated grading software, this instructor suggested that comments on recommended readings be invisible to fellow students by default. This would allow course staff to remove malicious content and notify the students. Another course instructor in this group complained ".. *unfortunately, students are often testing the limits of the honor code.*".

An Engineering instructor cited his time in the industry as having informed his singular emphasis on problem-solving in teaching graduate courses. He mentioned he favored creating and updating course assignments, projects, and exams without serious consultation from a primary textbook. This, according to him, discouraged students' use of online solution manuals, better assessed their progress with course milestones, and ensured they learned a precise set of Engineering skills without **undue cognitive burden**. Note the parallel with **AV**-faculty's rationale for the limited role of a 'Suggested Readings' RS in teaching courses for professional diplomas. About restricting students to a comment-only RS editorial role, he commented:

> "*It's not so much that I want the control. It's more that I don't think that should be there focus from an educational standpoint. I'm trying to get them to comprehend some pretty intense Engineering and design concepts. I consider it a burden for them to go out and find other resources. I really have to get them to practice a lot of things over and over.*"

As per table 4.4, several **V**-faculty suggested they might use the recommender system in **supporting research**, as opposed to assorting readings for teaching purposes. The frequently cited use-cases for such a recommender system were discovering research articles beyond Virginia Tech library-indexed databases, and discovering topic overlap with other research fields to inform literature reviews. It is worth noting that no teaching assistants or students favored a **V**-role for students, hinting at a strongly asymmetric editorial relationship between faculty and students. A more rigorous evaluation of this power asymmetry is left for future work.

4.4.2 Mapping Recommendation Needs

Recommender systems for education have fulfilled a variety of analytics tasks for the individual learner, interpretation and intervention, in-class and online [104]. However, there is a pressing need for the educational RS community to recognize platform-level changes in the domain. Recommender systems have to reckon with concerns of trust, efficacy, and interpretation *at scale*. The different faculty-preferred RS use-cases in Table 4.4 (teaching, professional development, research) begin to suggest for institutional support personnel a need for flexibility and personalization in the selection and targeting of recommendation tools towards faculty. As we observe in Study III later, these editorial modes ('editor', 'author', 'active viewer', and 'viewer') and their frequent rationales and use-cases are associated with status-quos of specific initial trust relationships between faculty, teaching assistants, and students. They can thus inform a variety of design choices (audiences, content, transparency, preference elicitation) in the design of an institutional, closed-loop recommendation strategy.

4.5 Limitations and Future Work

Our hypothesis tests and interviews in Study II reveal a spectrum of faculty's editorial trust intentions for students, ranging from conservative (students can view or rate recommended course readings) to egalitarian (students can perform recommendation authorship and editing tasks). Our preliminary study points to the link between editorial trust beliefs and intentions.

However, our study survey relies on single-item trust attribution questions, which limits the robustness of our findings. In Study III, we address these limitations by using a multi-target, multi-belief trust survey based in McKnight's typology [117] of trust beliefs (competence, benevolence, integrity).

4.6 Chapter Summary

In this chapter, we learn that faculty's allocation of editorial tasks to students in the context of an educational recommender system is linked to the trust placed in both students and the recommendation algorithm. We also discover that faculty express a range of preferred modes of editorial autonomy for students, ranging from editorship (typically in graduate STEM courses) to viewership (high-enrollment, undergraduate STEM courses). Faculty who prefer RS editorship and authorship for students typically believe in an instructor-in-the-loop content moderation model and a belief in creating initiatives for student engagement. Faculty who prefer RS viewership for students tend to mention the risks of inappropriate content (spam, misinformation), distraction from dictates of the curriculum, and a perceived lack of student engagement.

Chapter 5

Supporting Faculty's Delegation and Transparency Preferences

In our previous studies, we discover that faculty's preferred allocation of editorial tasks towards students for educational recommendation exists on a spectrum. This allocation ranges from *viewership*, where students can view suggested readings (and rate or comment on them, in a subset of cases) but not create or remove them, to *editorship* where students are actively involved in all or most authoring and feedback tasks. We also observe that the tendency to allocate editorial tasks to students is linked to perceptions of trust towards students. In Study III, we investigate if explanations, a transparency cue to describe stakeholder intent for recommendation users, can improve trust, increase delegation, and facilitate reflection across these editorial contexts. This helps UX researchers reflect on the right audiences, content, use-cases, and process transparency for educational recommendation. We expand on the design of Study II in two important ways. We include the RS algorithm as a stakeholder in the notion of editorial authority delegation, and examine three key trust beliefs (competence, benevolence, integrity) allocated between all stakeholders. This helps us probe the complexity of the trust relationships driving faculty's editorial intent.

img/studythree_overview.png

Figure 5.1: Study overview: relational attributes (trust perceptions, allocation of editorial tasks in recommendation), personal attributes (transparency affordances, trust propensity, expertise, leadership), stakeholders, and evaluation methods.

5.1 Study Objectives

Study III evolves the notion of editorial authority (**E-Auth**) from Study II into a multi-stakeholder, AI-integrative notion of delegation we refer to as **DelMo** ("delegation model"). It further investigates the distribution of trust beliefs of competence, benevolence, and integrity for our study stakeholders (faculty, TAs, students, RS algorithm). It examines our stakeholders' preferences for transparency in recommendation for TEL, their connections with delegation, and their ability to facilitate trust, reflection, and evolution in delegation preferences. The study concludes with a synthesis of best practices for the design of trustworthy recommender systems and effective institutional support, in view of faculty's LMS use and intent to delegate.

5.2 Study Design: Delegation, Trust, and Transparency

In this section, we introduce our study variables (visualized in figure 5.1) and primary research questions. The study variables include (a) preferred modes of delegation of editorial img/delmo1.png

Figure 5.2: The delegation mode (**DelMo**) from the reference point of one (or multiple) stakeholders is the sum of # tasks allocated to every other stakeholder, weighted by the # tasks allocated to all stakeholders. In this example, **DelMo** is depicted from the reference points of staff (left) and faculty (right). Recommendation tasks of *seed*, *veto*, *rate*, *comment*, *refresh* and *respond* are indicated by **S**, **V**, **R**, **C**, $\mathbf{R}_f$ and $\mathbf{R}_p$, and staff refers to faculty and teaching assistants considered together. Note that both of these **DelMo**s can be calculated for each individual respondent in our study.

tasks, (b) trust beliefs pertaining to RS stakeholders, and (c) transparency affordances. The research questions **RQs I - III**, in turn, investigate frequent patterns and linkages pertaining to these variables.

5.2.1 Modes of Editorial Delegation

The "delegation mode" of RS stakeholders (one or multiple) in set **REF**, notated as $\mathbf{DelMo}_{\mathbf{REF}}$, as the sum of # tasks allocated to every other stakeholder, weighted by the # tasks allocated to all stakeholders. Six recommendation tasks (*seed*, *veto*, *rate*, *comment*, *refresh*, *notify*, see Appendix D.2.2) are drawn from prior literature [78]. Equation 5.1 represents the delegation metric for all stakeholders in the set $\mathbf{REF} \subset \mathbf{U}$, where $\mathbf{U}$ is the set of all RS stakeholders.

$$\mathbf{DelMo_{REF}} = \frac{\sum_{i \in (U-REF)} N_i}{\sum_i N_i} \tag{5.1}$$

This simple metric helps us assess how every study participant allocates editorial tasks to faculty, TAs, students, and the RS algorithm. Figure 5.5, for instance, reveals that a disproportionately large share of recommendation vetoing in our study survey is assigned to faculty, but not to students. These study participants, therefore, favor an egalitarian mode of delegation for faculty, and a conservative mode of delegation for students.

Figure 5.2 visualizes the delegation metric $\mathbf{DelMo}_{\text{STAFF}}$ for **REF** = {Faculty, Teaching Assistants}, and $\mathbf{DelMo}_{\text{FACULTY}}$ for **REF** = {Faculty}. Here, **U** = {Faculty, Teaching Assistants, Students, RS Algorithm}. In this illustration, students and RS algorithm can perform two tasks each (*rate* and *comment* on recommended readings), TAs can perform four (*seed*, *rate*, *comment* and *respond*), whereas faculty can perform all six. Therefore, $\mathbf{DelMo}_{\text{STAFF}}$ is 4 divided by 14 (28%), and $\mathbf{DelMo}_{\text{FACULTY}}$ is 8 divided by 14 (57%). Intuitively, the delegation metric is a simple fraction of editorial tasks handed away from stakeholders in **REF**, such as staff and faculty in the example above. We identify respondents as 'Egalitarian' and 'Conservative' in the editorial process if their **DelMo** is above or below the population median, in that order. We focus on $\mathbf{DelMo}_{\text{STAFF}}$ for the remainder of our study, because it models how tasks are handed to stakeholders *without* an instructional prerogative (students, algorithm).

5.2.2 Trust Perceptions and Transparency Affordances

Our study examines three key trust beliefs (competence, benevolence, integrity) identified by McKnight et al. [117], about faculty, teaching assistants, students, and RS algorithm. In the context of our study, the *competence* of a trustee refers to their adequate knowledge of course topics and their ability to help curate course readings. *Benevolence* refers to their good intentions, and the tendency to appreciate and advance the interests of the truster. For instance, faculty's interests can include fulfilling the learning objectives of a course, meeting professional curriculum standards, managing large classes, and ensuring student safety and privacy. *Integrity* refers to the strength of the trustee's moral principles and attitudes. Trust relationships are assessed by the relative magnitudes of these three beliefs between any two stakeholders. This allows us to evaluate **asymmetry** in these relationships: instances where one stakeholder confers trust more than it assesses it for oneself (Section 5.4.2). We also contend that editorial intentions express the *intent to trust*. This is because trust intentions are typically understood as the willingness to assume some risk in interacting with a digital artifact. McKnight et al. [117] note that the customer of an online retail system indicates their trust by volunteering personal information or making a purchase. In higher education, a faculty member can delegate recommendation authoring or editing tasks, with risks like mishandling of student feedback, misinformation, spam, data loss, or absenteeism.

We evaluate three key RS transparency affordance groups: authorship cues (**AC**), item rationales (**IR**), and algorithm attributes (**AA**), each with three constituent types described in Table 5.1. **AC** represents transparency cues about the *process* of curating course materials, **AA** highlights RS features and capabilities, and **IR** emphasizes content-quality and use-quality cues. Figure 5.3 provides sample illustrations of each transparency affordance group, as presented to our study participants.

Three different types of explanations in the educational recommendation context are assessed in this study:

- Authorship Cues: '30% of your classmates found these useful'
- Item Rationales: 'further reading on binary search from week 2'

img/sample_explanations.png

Figure 5.3: Sample illustrations of transparency cues centered on recommendation *rationales*, *process*, and *features* provided to study participants, in that order: (Top, left) item rationales (**IR**), (top, right) authorship cues (**AC**), (bottom) algorithm attributes (**AA**)

Table 5.1: Key transparency affordance groups and types examined in our study, with examples.

Group	Type	Motifs, Topics, Examples
Authorship Cues	Source + target + topic	"X% of faculty found this helpful for students in explaining merge sort"
(**AC**)	Source + target	"X% of faculty found this helpful for students"
	Source	"X% of faculty found this helpful"
Item	Course deliverables	"HW1", "Final Project", "Extra Credit Quiz #2"
Rationales	Types + topics	"Research article", "Book chapter", "UX design"
(**IR**)	Time investment	"30 min read", "7 min skim", "Difficulty: medium"
Algorithm	Control methods	[Hyperlinks to RS setup: access, editing, notifications]
Attributes	Quality concerns	[Safety how-tos to prevent data loss, spam, misinformation]
(**AA**)	Time investment	[RS setup and maintenance time]

- Algorithm Attributes: algorithm capabilities, conventions and constraints

The study objectives are as follows:

(I) Assess if authorship cues are perceived as more effective than item rationales and algorithm attributes,

(II) Assess the distribution of initial trust relationships (competence, benevolence, integrity) between higher education stakeholders,

(III) Assess if the RS editorial trust intentions (allocations of seed, veto, rate, comment, refresh and notify tasks) expressed by stakeholders are linked to their mutual initial trust relationships,

(IV) Assess if RS transparency affordances (authorship cues, item rationales, algorithm attributes) are perceived to facilitate reflection and nudge editorial preferences,

(V) Assess how LMS platform contexts and editorial intent for educational recommender systems can jointly assist with the design of effective institutional support and establishing or reinforcing stakeholder trust

Table 5.2: Key counts for our study survey sample. Department counts (overall, STEM) appear in parentheses next to phase totals.

Stakeholder	All Participants # (Dept, STEM)	Authorship Cues (**AC**)	Item Rationales (**IR**)	Algorithm Attributes (**AA**)
All	501 (74, 53)	148	200	153
Faculty	29 (36, 26)	9	9	11
TAs	35 (21, 17)	9	7	19
Students	437 (43, 33)	130	184	123

5.3 Datasets and Methods

Study III was conducted in three steps (figure 5.1). We recruited participants for all three steps on a rolling basis during the Summer I, Summer II and Fall academic terms, between April 2022 and May of 2023. We leveraged convenience sampling and voluntary response sampling on departmental mailing lists and Facebook groups for recruitment purposes, and solicited participants' consent to participate in the interview stage in the concluding section of our study survey. The faculty members, students and teaching assistants who participated in our survey represent 74 departments in total. 53 of these 74 departments represent STEM disciplines of study.

The first step is a cross-department survey at Virginia Tech, with 501 participants (29 faculty, 35 teaching assistants, and 437 students). The survey covers trust beliefs, intent to delegate, transparency affordances, preference for automation, expertise, and leadership. The trust questions are based on McKnight's typology of trust beliefs (competence, benevolence, integrity) [117], with 9-point Likert-style response items (strongly disagree-strongly agree). Sections 2.2.2 and 5.2.2 dive deeper into these trust beliefs in the context of educational recommendation. Survey questions are outlined in Appendix D.2. We distinguish between (1) overall trust for a given cohort as % agreement in the cohort (% of individuals in the cohort with strong or partial agreement in all three trust beliefs), and (2) individual trust as average agreement level across beliefs. While we report both in Study III, we use the latter for our hypothesis tests to avoid potential biases in the first from bucketing and aggregation. We perform hypothesis tests to assess the relationships between our study variables using one-way ANOVA (F-test). Table 5.2 details key attributes of respondents in the survey.

The second step is the analysis of transcripts and notes from follow-up semi-structured interviews with 18 participants (6 faculty, 3 teaching assistants, 9 students) and responses to open-format questions from the previous step. Appendix E.2 outlines the interview questions. We perform thematic analysis and affinity diagramming [69] on survey responses and interview transcripts to understand respondents' frequent rationales for their preferred delegation modes, trust relationships, and transparency affordances (highlighted in bold in Section 5). Figure 5.4 illustrates this process. These analyses also help us synthesize design guidelines for trustworthy educational recommendation and institutional support.

The third step is a small-scale survey of Virginia Tech DoIT/TLOS staff members (N=13) on their policy preferences for RS audiences, content, and transparency. Appendix D.3 describes the survey questions used in this step.

5.4 Findings

Figure 5.5 outlines key distinctions in the way faculty, teaching assistants, and students allocate recommendation tasks. Fewer than half of survey respondents assign recommendation veto-ing to teaching assistants, and fewer than a quarter to students and the RS algorithm. There is considerable consensus among all stakeholders on near-exclusive article vetoing by course instructors, and in a small subset of cases, by course staff. Teaching assistants often allocate seed and veto tasks near-exclusively to instructors, however, they express greater intention to delegate relative to instructors for the rating and commenting tasks. On average, img/qual_process.png

Figure 5.4: Thematic analysis of Study III survey comments and interview transcripts

img/deleg_alt3.png

Figure 5.5: Initial task allocations preferred by stakeholders (instructors, teaching assistants, students) in our study. For instance, 93% of faculty think faculty should seed recommendations, whereas 31% think students should. 90% of faculty assign the RS veto task to faculty, whereas 24% assign it to teaching assistants and only 3% to students or the RS algorithm.

img/trust_alt.png

Figure 5.6: Trust beliefs of competence, benevolence, and integrity about all stakeholders (% agreement). For instance, all survey respondents express the least trust in the RS algorithm relative to other stakeholders, and student benevolence is ranked lower than their competence and integrity.

img/delmoxtrust.png

Figure 5.7: Trust levels allocated to Study III stakeholders (faculty, teaching assistants, students, RS algorithm) as a function of DelMo (conservative and egalitarian modes).

students have higher preference for delegation relative to course staff but their allocations can be equally selective. Finally, all respondents perceive content tags and social proof as more useful, effective and trustworthy, relative to algorithm primers. Table 5.3 outlines the results from key hypotheses pertaining to study **RQs I-III**.

Delegation (RQI, H1-H3)

We learn that faculty tend to delegate editorial tasks less than teaching assistants and students ($F = 6.6, p = 0.01$). Teaching assistants, however, do not delegate editorial tasks less than students ($F = 0.36, p = 0.54$). This preference for delegation is positively correlated with perceived expertise in recommendation algorithms, and graduate students express higher preference for delegation relative to undergraduate students.

Table 5.3: Key results from hypothesis tests about the relationship between delegation and trust perceptions. 'Course staff' refers to faculty and teaching assistants considered together. For instance, according to **H1**, faculty have a lower tendency to delegate RS editorial tasks compared to students and TAs. However, according to **H3**, TAs do not prefer to delegate less than students do.

Research Questions	Hypotheses	$\|F\|$
(**RQI**) Delegation	**H1**: Faculty prefer to delegate less than TAs and students do.	6.6*
	H2: Preference for RS delegation increases with perceived RS expertise.	6.1*
	H3: TAs prefer to delegate less than students do.	0.36
(**RQII**) Trust Perceptions	**H4**: Faculty prefer to trust stakeholders less than TAs/students do.	18.6**
	H5: Faculty tend to trust RS algorithm less than TAs/students do.	31.6**
	H6: Faculty believe in student competence less than TAs/students do.	14.3**
	H7: Faculty believe in student benevolence less than TAs/students do.	31.7**
	H8: Faculty believe in student integrity less than TAs/students do.	1.1
(**RQIII**) Delegation & Trust	**H9**: Delegation to students/RS algorithm increases with increase in trust placed in students/RS algorithm.	8.36**
	H10: Delegation increases with propensity to trust.	0.15

*stat. signif., $\alpha = 0.05, p <= \alpha \wedge p > 1e{-}10, ** \, p < 1e{-}10$

Trust beliefs (RQII, H4-H8)

Faculty tend to trust stakeholders less than teaching assistants and students, both overall ($F = 18.6, p < 0.01$) and in competence ($F = 14.3, p < 0.01$) and benevolence ($F = 31.7, p < 0.01$) beliefs regarding students. They do not have a significant gap in their assessment of student integrity than students and teaching assistants do. Figure 5.5 illustrates that student benevolence, in particular, is assessed to be about half that of other stakeholders. Finally, faculty are typically more hesitant about algorithmic agency in the short-term, although many are open to automated recommendations in the longer-run.

Delegation and stakeholder trust (RQIII, H9-H10)

Higher trust in students and the RS algorithm trust is linked to higher tendency to delegate editorial tasks in their favor ($F = 8.36, p < 0.01$) for all stakeholders. Individual propensity to trust, however, is not directly linked to the tendency to delegate. Figure 5.7 illustrates the relationship between trust allocations to Study III stakeholders as a function of editorial modes (conservative and egalitarian). Egalitarians tend to have higher trust for students and the RS algorithm relative to conservatives. Table 5.4 identifies the permutations, or *arrangements* of editorial roles most commonly preferred by faculty members. The top two (accounting for 34% of the faculty members) are conservative towards students and the RS algorithm, whereas the next three (accounting for 23% of the faculty members) are egalitarian. We review the frequent rationales and motivations for these editorial contexts as follows. Note that in this section, primary themes from our content analysis appear in bold text, and stakeholder identifiers (**S**, **TA**, etc.) appear alongside key quotes pertaining to the primary themes.

5.4.1 Delegation Preferences: Conservative, Egalitarian

"The Conservatives": role-preserving mode

Study participants who favor this mode of delegation assign recommendation rating and commenting tasks to students and RS algorithm, and retain article seed, veto, refresh and respond tasks for faculty or teaching assistants. According to table 5.4, a total of 34% of faculty and 20% of students favor the two conservative editorial arrangements at the course level. The first allocates faculty as the sole "Editor" and the rest of the stakeholders as "Active Viewers". The second allocates faculty as the sole editor, but allows teaching assistants to author recommended readings. Both arrangements restrict article vetoing to course faculty. Our content analysis reveals that stakeholders cite **instructor prerogatives and competence** as the most common reason for their preferred delegation modes (table 5.4). They argue that it is crucial for faculty to maintain near-exclusive authorship privileges for their courses, because they are responsible for decisions on course curricula. The second most frequently-cited rationale for limiting RS authorship to course staff is presumed **authorship responsibilities** and burdens (algorithm supervision, removal of malicious content, managing feedback at scale).

> "*Control over their courses is a key faculty responsibility. It is also important for academic freedom. Only faculty should control their syllabus, from top to bottom.*" [**F16**]

> "*Instructors are responsible for curating what material gets included as part of the course. A recommendation system influenced by TAs and students will, in my opinion, result in the instructor having to spend time making sure the recommendations are appropriate and fit with the course objectives.*" [**F7**]

Stakeholders in this group frequently recognize the utility and **need of student feedback**, but they simultaneously cite **student incompetence** (incomplete knowledge of course topics), thus restricting their editorial tasks to rating and commenting. They also recognize the need for teaching assistants to bridge the disconnect of knowledge and expectations between faculty and students. We explore these themes further in Sections 5.2.1 and 5.2.2.

> "*I think students should have an opinion on whether they liked the reading or not, but not the power to choose it because there are so many students in a class, and it might be difficult to narrow down readings.*" [**TA13**]

> "*(The instructor's) feedback shouldn't be overruled/preempted by an algorithm, especially one that's likely inside a black box. TA's opinions should be given consideration by both the instructor and the algorithm, perhaps by moving their comments on a reading higher or weighting their 'up-vote' more when ranking.* " [**S24**]

Table 5.4: The top-5 most frequent editorial divisions-of-labor, as favored by faculty members in our dataset. **E**, **A**, **AV** and **V** refer to roles of "Editor", "Author", "Active Viewer" and "Viewer", respectively, while 'F', 'TA', 'S', and 'Alg' refer to our four study stakeholders: faculty, teaching assistants, students, and RS algorithm, respectively. For example, the most faculty-preferred editorial arrangement assigns every stakeholder except the faculty an "Active Viewer" role in curating course readings.

Assigned Roles (F, TA, S, Alg)	Delegation Mode (**DelMo** $_{\textbf{STAFF}}$)	% F	% S	Primary Themes
E, **AV**, **AV**, **AV**	Conservative	21.7	5.3	(1) Instructor prerogatives,
E, **A**, **AV**, **AV**	Conservative	12.82	15.1	(2) authorship responsibilities,
				(3) need for student feedback,
				(4) student incompetence,
				(5) staff benevolence
E, **A**, **A**, **AV**	Egalitarian	8.9	9.1	(1) Need for student feedback,
E, **E**, **A**, **A**	Egalitarian	7.7	7.2	(2) authorship responsibilities,
E, **E**, **A**, **AV**	Egalitarian	6.4	7.3	(3) staff competence

"The Egalitarians": role-sharing mode

Study participants in favor an egalitarian mode of delegation assign recommendation seed, veto, refresh, and respond tasks to students and RS algorithm. Table 5.4 highlights three egalitarian editorial arrangements preferred by 23% each of faculty and students, overall. The most frequently cited rationale for this preference is the **need for student feedback**, that is, students' ability to express to course faculty and teaching assistants how useful and effective the curated readings are towards their perceived academic needs and objectives. Course faculty in this group often acknowledge that student voices should be incorporated in the curation of course materials, as a vehicle to promote learning, discussion, transparency, and overall student engagement and autonomy. Stakeholders acknowledge the **authorship burdens** associated with maintaining a recommendation engine for the course, especially at scale and in the short-term, and encourage course staff-led supervision of the RS algorithm to minimize the possibility of spam, abuse, and bias in recommended course materials.

"*For the algorithm to learn each course best, it'd be ideal if all participating members (faculty, TAs, and students) were able to contribute to content, like and dislike, and comment on items.*" [**F5**]

"*Open and reciprocal engagement for students with the materials and even the class structure is foundational to my pedagogical approach.*" [**F15**]

"*For students submitting recommended readings, perhaps we can have the professors validate and approve all links provided by students, or have a 'student suggestions' tab where students can freely submit their own readings. The professors and TAs can move readings from the students' tab to the main class tab if the reading is truly useful.*" [**S29**]

5.4.2 Trust Perceptions: Competence, Benevolence, Integrity

In this section, we take a closer look at key constituent beliefs of RS stakeholder trust. We highlight the commonalities and disparities of perceived competence, benevolence and integrity between stakeholders, and their frequently cited rationales. In figure 5.8, the areas contained within each triangle visualize the overall trust received from a source stakeholder (truster) to a target stakeholder (trustee). For instance, faculty, teaching assistants, students, and the recommendation algorithm receive diminishing degrees of overall trust by all sources, in that order. Table 5.5 highlights the asymmetric exchange of trust between stakeholder pairs.

img/trust_composition.png

Figure 5.8: Trust perceptions held by a (truster) stakeholder towards a (trustee) stakeholder. Trusters include faculty (left), teaching assistants (middle), and students (right). Perceptions include beliefs about stakeholder competence, benevolence, and integrity. For instance, about 90% of faculty members express a belief in faculty's overall competence (left, orange radii). However, only about 40% of faculty members believe in RS algorithm's competence (left, brown radii). It is worth noting that if all three trust beliefs (competence, benevolence, integrity) are ranked equally by the truster, the triangles appear balanced and equilateral. For instance, faculty members rank the benevolence of students as significantly lower than their integrity, so the green radii in figure 5.8 (left) lack balance. Compare this, for instance, to the balanced yellow radii corresponding to the TAs. Table 5.5 quantifies these relationships.

Competence Perceptions

All study participants express comparably high degrees of belief in faculty's competence. According to figure 5.8, about 87%, 88% and 92% of faculty, teaching assistants and students, respectively, agree that course instructors are competent overall. This belief is frequently attributed to **perceived expertise** (in-depth knowledge of course objectives, content, or outcomes), and **experience** (length of teaching and research career, historical involvement with the teaching and design of a given course). A smaller subset of stakeholders, especially students, frame the expectation of competence in terms of **prerogative** (rights and privileges implied by faculty's job role), and faith in institutional hiring practices.

"*Instructors have a better grasp on learning objectives and outcomes and (rec-*

> *ommended) reading tie-ins.*" [**S13**]

> "*Instructors should have a firm grasp of the course, ... an ability to understand what material a student needs to learn in order to be successful in the course.*" [**TA2**]

> "*People in power that teach my classes are automatically trustworthy in my eyes. They have worked very hard to be where they are, and that's why I trust their judgement.*" [**S45**]

Teaching assistants are perceived by all study participants as competent for all the aforementioned reasons, albeit not as frequently as faculty are. About 77%, 81% and 83% each of faculty, teaching assistants and students rank teaching assistants as competent. Stakeholders attribute this difference of perceptions to limited **expertise** of TAs relative to instructors (including lack of relevant prior coursework) and **departmental factors** (limited supply, poor skill-match). Conversely, their prior **experience** with the course is often cited as a relative strength, especially in their ability to identify and respond to student needs. We explore this further in Section 5.2.2.

> "*As a full professor and course designer, in my experience, TAs assigned to a course often come from backgrounds that do not include the subject matter of my course. They do what they have to, as best as they can, but often make mistakes if left to grade open-ended questions.*" [**F45**]

> "*Teaching assistants have less knowledge about the course material but often know more helpful sources as they have taken the course.*" [**S14**]

Table 5.5: Trust asymmetries by belief type, evaluated using group differences (F-test) between trust conferred and self-assessed for stakeholder pairs (leftmost column). Statistically significant effects ($F > F_{crit}, p < 0.05$) are described in bold text. For instance, faculty rank student competence and benevolence significantly lower than students rank their own competence and benevolence.

Stakeholder Pair	Competence (F, p)	Benevolence (F, p)	Integrity (F, p)
Faculty -> Students	**11.4, <0.01**	**12.1, <0.01**	0.11, 0.74
Faculty -> TAs	0.33, 0.56	0.05, 0.8	1.5, 0.21
TAs -> Faculty	0.01, 0.89	0.27, 0.6	1.1, 0.29
TAs -> Students	0.63, 0.42	0.94, 0.33	0.71, 0.39
Students -> Faculty	1.65, 0.19	**5.6, 0.01**	2.3, 0.12
Students -> TAs	0.1, 0.75	**5.7, 0.01**	0.003, 0.95

> "*Teaching assistants take initiative and direction from the teacher, thus they have shared values. The instructor has a greater depth of knowledge though.*" [**S17**]

According to **H6**, faculty have significantly lesser faith in student competence than teaching assistants and students do. 61.5% of faculty rank students as competent, whereas 74% and 80% of TAs and students rank students as competent, in that order. Study respondents often frame this perception as self-evident given their status as a student, and expect considerable variation and uncertainty in their knowledge of course topics, and the ability to rank and curate course readings. However, the asymmetry captured in table 5.5 likely points to a deeper disconnect between faculty and students in assessment of student competence and benevolence.

> "*I sense a disconnect between student learning goals and instructor teaching objectives. The knowledge gap between students and instructors adds to the disconnect, especially for undergraduate students. It's more pronounced for large courses.*" [**F6**]

> "*Students, by definition, rarely have the background or the experience to judge the course materials, the course value, or the quality of the teaching.*" [**F45**]

> "*I perceive that students are the least engaged and empowered, so they'll be less effective in curating (the recommended readings).*" [**TA5**]

Study participants are similarly divided on the competence of the recommendation algorithm. 39.7% of faculty perceive the algorithm as competent, compared to 57% of TAs and 70% of students. While there is little evidence of a dispositional distrust of the recommendation algorithm, faculty members frequently express concerns about its potential **inefficacy** (inaccuracy, poor understanding of student or curriculum needs) and **bias** (racial, cultural, filter bubbles, echo chambers). These concerns are especially expressed for the short-term, when the recommendation algorithm is not presumed to have incorporated any human feedback. All stakeholders emphasize the need of human editorial supervision of the recommended readings, either proactive (mandatory review of recommendations before publishing) or reactive (editing or removal of content on a need-basis).

> "*I'm very reluctant to enable an algorithm for my history courses. The likelihood of an algorithm to incorporate something that is technically relevant but intensely partisan or factually inaccurate is very high.*" [**F27**]

> "*Our course content is well-established by our professional discipline and licensure requirements so the content would be clear to the algorithm. I'm not sure if the algorithm could make decisions about the developmental readiness of students, though.*" [**F24**]

Benevolence Perceptions

Participants in our study perceive faculty as benevolent most frequently among all stakeholders, albeit with some disagreement on the precise set of **student needs** which deserve their attention. 85% of students rank faculty as benevolent, when 69% of teaching assistants and 73% of faculty do the same. Faculty and teaching assistants often acknowledge that **learning outcomes** of the course are their priority, and students' interests in the domain might not be easily identifiable or incorporated in the course design.

> "*Do you mean students' psychosocial needs or learning goal needs? My primary concern is the latter.*" [**F24**]

> "*I would assume course instructors are concerned with transferring domain knowledge, and less concerned with the more nuanced needs of students, like time, accessibility, interestingness of the reading.*" [**S80**]

77% of students rank teaching assistants as benevolent, compared to 62% and 64% of faculty and teaching assistants, in that order. Students frequently cite the ability of TAs to engage with them one-on-one, and to appreciate student needs better than faculty do. Table 5.5 also reveals that students tend to make optimistic assessments of staff benevolence relative to staff's self-assessments on the matter.

> "*TAs are still students.. and therefore may empathize more strongly with students than the teachers do.*" [**S53**]

> "*TAs' interests are more likely to align with my own because they are current or recent students themselves. They have a closer connection and strong memory*

> *of what type of things they found to be engaging or helpful in understanding the material and potentially sparking increased interest in the student.*" [**S41**]

We find a huge variation in the benevolence beliefs regarding students and the recommendation algorithm. Only about 35% of faculty rank students as benevolent, compared to 60% of students, the difference also evidenced by hypothesis **H7**. This particular deficit of trust appears frequently in high-enrollment, undergraduate classrooms. Faculty members in these contexts complain about a general **lack of student motivation** to complete course obligations in a timely fashion, and **outcome bias** (exclusive focus on grade attainment). Algorithms are ranked as benevolent by about 23%, 48% and 58% of faculty, TAs, and students, in that order, for concerns mirroring the algorithm competence concerns in Section 5.2.2 (bias, inefficacy, need for human supervision and feedback).

> "*I'm sure the algorithm will generate reasonable recommendations. However, I can't motivate my students to do required work, much less recommended reading.*" [**F21**]

> "*Students are wildcards. Some might be interested in exploring topics further and providing good suggestions, while others may just be trying to mess with you or do not really know what they are talking about.*" [**S10**]

> "*Most of my experience with algorithms is more tailored towards achieving business outcomes than my own needs. I don't have a ton of confidence that we would be able to develop recommendation algorithms in the short term that would adequately reflect the needs of students.*" [**S81**]

Integrity Perceptions

Beliefs about integrity of faculty, teaching assistants, and students feature ample consensus, with 89%, 80% and 76% of the survey respondents in agreement with the integrity perception for these three groups, respectively. One key exception to this consensus is in the case of large, multi-section, undergraduate courses where faculty members cite **honor code violations** (cheating, plagiarism, complicity) as rationale for lower relative perceptions of student integrity.

The recommendation algorithm lags behind the human stakeholders substantially, with about 25%, 50%, and 57% of faculty, TAs, and students agreeing to its integrity, respectively. Stakeholders frequently acknowledge that these integrity concerns, in fact, represent a **lack of knowledge** about, and the **perceived limitations** of, the processes of search, ranking and personalization encoded in the algorithm. Faculty members in our survey especially raise concerns of systemic racial and cultural bias in black-box recommendation algorithms. They overwhelmingly recommend human-in-the-loop governance, transparency, and accountability of recommendation processes, to limit the scope of these algorithmic biases.

> "*Algorithms seem biased and geared towards data harvesting.*" [**S29**]

> "*Students live in a media environment saturated by algorithmically-generated recommendations. These algorithms are responsible for a host of biases and general poor performance. My job as an instructor is to defeat those algorithms and provide a curated space for ideas open to contestation.*" [**F4**]

img/tafs4.png

Figure 5.9: Overall perceived trustworthiness, usefulness, and persuasiveness of the constituent explanation types in each of the three transparency affordance groups in our study: item rationales (left), authorship cues (middle), and algorithm attributes (right). For instance, 79% of all stakeholders consider item rationales about course deliverables (homework assignments, project milestones) trustworthy, while 87% of the stakeholders consider authorship cues signaling course faculty as curators as trustworthy.

Relationship with Delegation

Hypothesis **H9** illustrates a key relationship between trust perceptions and preferred delegation of RS editorial tasks. Higher instances of trust beliefs for students and the RS algorithm imply higher likelihood of an egalitarian mode of delegation for them. When broken down by specific trust beliefs, this relationship is the strongest for perceived benevolence ($F = 6.9, p < 0.01$) and competence ($F = 4.14, p = 0.04$) of students/RS algorithm. Perceived integrity is not linked to delegation modes ($F = 0.04, p = 0.83$). This illustrates that the significant variation and lack of consensus in perceived benevolence of students and the RS algorithm (per our observations in Section 5.2.2) is consequential for the degree to which course staff plans to allocate editorial responsibility to these stakeholders.

5.4.3 Transparency Preferences: Authorship, Rationales, Algorithm-Awareness

Figure 5.9 illustrates faculty and students' assessment of key transparency affordance groups in Phase 2 of our study. 87% of the respondents ranked specific authorship cues (especially, decision source, target, and topic) as trustworthy, compared to 79% for the most trustworthy IR (course deliverables attributed to the recommended item). Authorship cues typically exceed item rationales and algorithm attributes in trust, persuasiveness, and utility. They are also ranked by our study stakeholders the most likely to nudge their preferred overall delegation mode. Trust in all three affordance groups is also consistently correlated with beliefs of student and algorithm competence, benevolence, and integrity. Two interesting patterns emerge when we examine these preferences by stakeholders and preferred delegation modes. One, faculty members tend to rank all affordance groups as less trustworthy relative to students, but this disagreement is most pronounced for authorship cues ($F = 23.4*, p < 0.01$), followed by algorithm attributes ($F = 9.2*, p < 0.01$) and item rationales ($F = 8.6*, p < 0.01$). This is in agreement with hypotheses **H4** and **H5** (table 5.3), as well as our observations on competence and benevolence gaps in Section 5.2. Authorship cues are often subject to serious scrutiny by faculty members to ensure they emphasize the role of the primary instructor (as opposed to department colleagues or the instructional design team) in curating and maintaining the recommended course materials. Two, the egalitarian group ranks authorship cues as trustworthy significantly more than the conservative group ($F = 9.47, p < 0.01$), but is ambivalent about item rationales ($F = 0.51, p = 0.47$) and algorithm attributes ($F = 0.03, p = 0.85$). These observations are in agreement with the motives of conservatives and egalitarian modes we outline in Section 5.1, especially, instructors' prerogatives and the need for student feedback. This also underscores the fact that our framing of delegation can validate insights into key editorial trust relationships and deficits,

Table 5.6: Frequency of trust perceptions about RS transparency affordances, overall and by delegation mode ("Conservative", "Egalitarian"), along with their perceived potential to inspire reflection and nudge these modes. For instance, 86.2% of all survey respondents, 60% of conservatives, and 40% of egalitarians rank authorship cues (**AC**) as most likely to inspire trust in the recommender system. Authorship cues are ranked higher in perceived trust relative to item rationales and algorithm attributes.

Affordance Group	% Trust (All, Cons, Egal)	% Reflect (All, Cons, Egal)	% Nudge (All, Cons, Egal)
Authorship Cues (**AC**)	86.2, 60, 39.9	77, 77.7, 75.8	61.4, 58.8, 65.5
Item Rationales (**IR**)	77.3, 68.9, 31	76.6, 76.5, 76.9	50.6, 49.5, 53.8
Algorithm Attributes (**AA**)	75.4, 69.9, 30	75.8, 74.5, 78.7	54.9, 51.8, 61.7

as well as their implications for objective aspects of an educational recommender system. These observations set the stage for us to explore the affordance-specific barriers to trust expected by Phase 2 stakeholders, and the associated opportunities for design. We briefly review these in the next section.

Frequent Barriers to Trust and Nudging

Table 5.7 reviews the most frequent barriers mentioned by our Phase 2 study participants. For authorship cues (AC), two primary barriers to emerge from the analysis were (1) **unreliable outside sources**, such as instructors affiliated with the department, not directly with the course, and (2) **lack of rationales**, such as an in-depth justification of why the judgement of fellow students was relevant to the recommended reading. For item rationales (IR), stakeholders complained about **potential inaccuracies** in read time and skim time estimates, and cautioned against reader discouragement. For algorithm attributes (AA), stakeholders complained about **information clutter** and **inability to verify** or trust the efficacy of solutions to quality issues.

Finally, participants suggested three broad remedies to address the aforementioned trust

Table 5.7: Frequent barriers to trust in key RS transparency affordances: authorship cues (**AC**), item rationales (**IR**), algorithm attributes (**AA**). For instance, our survey respondents identify a lack of rationale (why) and process (how) descriptions among frequent trust barriers for **AC**.

Affordance Group	Affordance Type	Frequent Barriers to Trust
Authorship Cues (**AC**)	Source + target + topic	Untrustworthy outside source (19%)
	Source + target	No rationale (44%)
		Untrustworthy source (10%)
	Source	No rationale (47%)
		Untrustworthy source (22%)
Item Rationales (**IR**)	Course deliverables	Lack of detail (17%)
	Types + topics	No rationale (38%)
		Too much info (15%)
	Time investment	Too much info (20%)
		inaccurate read times (5%)
Algorithm Attributes (**AA**)	Control methods	Too much info (34%)
		Lack of clarity (22%)
	Quality concerns	Lack of clarity (24%)
		Ineffective solutions (19%)
	Time investment	Inaccurate time estimates (25%)
		Lack of detail (25%)

barriers. First of these is **hybrid explanation paradigms**, combining IR (especially the course deliverables relevant to recommended reading) and AC (especially the *source + target + topic* motif). The second remedy is **highlighting source reliability** to establish user safety and privacy. The third and final remedy is to **avoid evidence bloat** by taking focus away from metrics susceptible to variation across courses, such as article read time and difficulty level.

img/policy_prefs.png

Figure 5.10: RS audiences, content and transparency as a function of faculty's Canvas use and delegation. Survey respondents are members of institutional support staff at Virginia Tech (IT, TLOS).

5.5 Contributions and Implications for Practice

In this section, we synthesize design guidelines to help UX researchers improve the trustworthiness and transparency of educational recommender systems. We also provide reflections for institutional support personnel on improving process efficacy in faculty outreach. We draw on our work on LMS platform contexts (Chapter 3), faculty's editorial intent (Chapter 4, Chapter 5), and trust relationships (Chapter 5) to discuss the corresponding design choices and tradeoffs.

5.5.1 Designing for Transparency and Trust in Educational Recommendation

Figure 5.10 describes how Virginia Tech IT and TLOS staff think about the personalization of recommendation audiences, content, and transparency based on faculty's Canvas use and delegation preferences (see Appendix D.3 for survey questions). They tend to recommend faculty-facing professional development content for low and medium LMS-use cohorts, and thorough explanations of the recommendation algorithm for faculty with low or medium LMS-use and conservative editorial intent. We synthesize our findings from this study into design guidelines as follows:

Audiences and Content: Acknowledge trust deficits.

In figure 5.8, we observe that recommendation stakeholders hold complex and sometimes asymmetric trust beliefs about each other. For example, faculty rank student competence and benevolence significantly lower than students rank their own competence and benevolence. 90% of faculty members express a belief in faculty's overall competence, while only 40% believe in RS algorithm's competence. These trust deficits are also linked to faculty's varied editorial preferences. Selecting the right audiences for educational recommendation should begin by acknowledging and managing these trust deficits. Faculty-facing recommender systems on the LMS should prioritize professional development, technology literacy, and new tech evangelism for faculty with medium LMS use and conservative editorial intent. Student-facing recommender systems should, in comparison, focus on faculty with high LMS use and egalitarian editorial intent. Collaborative curation of learning materials may encourage engagement and self-efficacy in students. However, the burden of content moderation and perceived outcome bias in students render these completely infeasible for many instructors with conservative editorial intent.

Transparency: Clearly articulate who is voting on the recommended readings.

Explanations based in social proof and authorship ("70% of students found this useful in understanding merge sort"), should clearly indicate who is voting on the utility of the recommended readings, how many are affiliated with the specific course using the recommender system, what their relevant expertise is, and what they found useful and persuasive. Users with all kinds of editorial intent strongly support the idea of situating the social proof within the classroom, and prioritizing the votes of faculty, students and teaching assistants affiliated with the class, as opposed to department peers or the IT instructional design team. Some study participants suggest that, instead of referring to peers from an unspecified department ("70% of instructors in your department"), instructors who recommend these readings should include their names and affiliations. This is because many courses are likely to enroll students from a variety of majors and pathways. Study III survey reveals a range of opinions on whose votes should be reflected in these explanations, consistent with the trust relationships we observe in Section 5.4.2. Course staff's expertise, and their prerogative to curate course content and grade assessments is frequently acknowledged, and faculty-sourced explanations are ranked among the highest in usefulness and trust (figure 5.9). While many consider student competence and trustworthiness to be lacking in completing or effectively evaluating the recommended readings, the persuasiveness of student-sourced votes is ranked highly. Some participants favor in-depth social proofs, such as voting percentages student reviews to help them make sense of the common challenges with each recommended reading. But, it is unclear if this additional information is desired by all stakeholders equally, regardless of their editorial intent (Section 5.4.1).

Transparency: Balance accuracy and conciseness in providing additional information about the algorithm and the recommended items.

Algorithm primers and recommended item rationales should carefully balance the competing needs for accuracy and conciseness. Course instructors and teaching assistants tend to ask for additional information about the RS algorithm's setup time, safety, and privacy more than students do. However, as we learn in current and previous studies, our study stakeholders' technology self-efficacy and appetite for new information are often on a spectrum. Too much information about these topics in a single dropdown menu generated complaints of visual clutter, cognitive overload, and lack of answers to common troubleshooting queries. Study participants suggest different ways to manage this information, including (1) an FAQ section with simple, non-technical language and short videos to reach a broad on-campus audience, and (2) multiple tabs to manage advanced information about the RS algorithm, with hyperlinks in the FAQ section, or a separate landing page, for those interested in this information. Estimates about the read time and difficulty of a recommended reading also generated concerns, primarily from students, about accuracy, unclear sourcing, potential de-motivating effects, and an overall lack of trust. Students appear more sympathetic to usefulness ratings (and full-length reviews, in a subset of cases) for these readings by staff and fellow students. A subset of study participants also favored the idea of estimated ranges for algorithm setup time and read time, albeit with clear indication of data sources to bolster their perceived credibility and trust. Clearly identifying the subject matter, topic area, or concept a recommended item can help students learn is often noted by our study participants as a powerful tool to increase RS efficacy and trust. Item rationales containing source, target and topic information together are ranked the most useful across explanation groups. However, as we note in Section 5.4, evidence bloat and cognitive overload can undermine trust in the recommendation process, and the volume, order and complexity of information presented in algorithm primers should be carefully managed to deal with these challenges.

Transparency: Establish periodic review of algorithmic decisions, and clarify support processes.

Algorithmic decisions about curated readings should undergo periodic reviews. Our learnings in Sections 4.4 and 5.4 reveal a range of opinions on who should be allocated these reviews (faculty, teaching assistants, students, institutional support personnel), and what their precise editorial powers should be (editor, author, active viewer, viewer). Conservative editorial intent corresponds with a preference for instructors and teaching assistants to edit and review the recommended readings, whereas egalitarian intent allows for student input in the process. Our analyses of stakeholder trust note that RS algorithm typically ranks last in trust relative to other stakeholders across belief types (Section 5.4.3), and is likely to be trusted more long-term after human review has established its overall efficacy in the short-term (Section 4.3). Instructors with both kinds of editorial intent want additional information about algorithmic curation, such as the relevance and trustworthiness of recommendations, data sources, curation logic, role management, risks of misinformation and spam, and security of student data.

Many of our study participants express concerns about low literacy with Canvas services, and acknowledge the utility of hands-on technical support and troubleshooting. The explanations should, therefore, clearly communicate how to get IT help in a variety of ways (IT helpdesk tickets, consults, wikis, on-demand digital skills coursework). The notion of periodic check-ins with support staff for algorithm setup generated some concern about the time expectation for these check-ins. Nonetheless, many noted that periodic notifications about setting up the recommender system on Canvas, similar to notifications for teacher evaluations and grading deadlines, may increase the overall use frequency of the system.

5.5.2 Designing for Process Efficacy and Trust in Institutional Support

Our analyses in the previous chapters reveal that (1) degree of faculty's LMS use is potentially linked to course modality, participation, logistics and outcomes, and (2) instructors for conservative and egalitarian course arrangements are likely to have different degrees of overall stakeholder trust and different preferred recommendation use-cases. Designing a vision of instructional support for these cohorts, we argue, requires a respectful consideration of their preferred use-cases and initial trust beliefs. The previous section reflected on how our knowledge of faculty's LMS use and editorial intent can drive the design of recommender systems for teaching, learning and outreach. In this section, we describe how these recommendation use-cases can assist IT personnel (designers, developers, project managers, IT leadership) in enabling the design vision we outlined in the beginning of this dissertation (figure 1.1). This vision comprises of the use of learning management system and trust-aware recommender systems to scale, personalize and iterate Virginia Tech DoIT's instructional support efforts, as follows:

Build coalitions and track relationships Faculty's engagement with institutional support efforts in the past is a crucial evidence point for their willingness to engage in future initiatives. IT personnel should invest in the capacity of tracking faculty's responsiveness to key outreach efforts (departmental mailing lists, online discussion groups, LMS announcements, social media). This information is often not readily available, or the datasets for different media are owned and managed by different teams within the IT organization. Bringing these datasets together and analyzing the effectiveness of these outreach media for different faculty cohorts can provide a nuanced look at faculty's sources of disengagement or mistrust. These cohorts can be defined in many ways, including using faculty's LMS use

(Chapter 3) and their editorial intent (Chapters 4 and 5).

We also recommend a revamping of standard-issue needs analysis surveys used by IT personnel to assess faculty's interest in new technologies such as blockchain or generative AI. These surveys rely on self-reports of faculty's competencies, knowledge gaps, and need for additional information and digital skills with regard to these new technologies. We recommend (a) complementing these self-reports with signals of LMS platform use (DOU) and editorial intent (DelMo), and (b) personalizing the needs analysis surveys to include questions on faculty-perceived potential increase in sensemaking, productivity and collaboration. These questions should use benchmark tasks relevant to faculty's existing technology use workflows, pedagogical approaches, and learning objectives from courses they have taught in the past. These signals can paint a richer portrait of faculty's adoption journeys, which can then be appropriately mapped to engagement nudges.

Acknowledge structural barriers and stakeholder bottom-lines. The design of technology artifacts and initiatives to support decision-making for university administrators and faculty, such as learning data analytics dashboards and causal inference tools, should recognize the distinct work roles, constraints, and success metrics of all stakeholders. For example, tenure-track research faculty at research-charter institutions may be allocated a division of duties (research, teaching, service) with a disproportionate share for research. Artifacts to help evaluate teaching quality, build digital skills, boost student productivity and identify struggling students, may struggle with faculty uptake in this environment. For IT personnel designing faculty development initiatives, this elevates the role of incentives (grants, assets, badges, credentials) to help increase participation. This also creates an effective upper limit on the work expected of faculty (number of hours spent on knowledge transfer, assignments, meetings) in these programs. Finally, navigating stakeholder trust relations is a necessary component of managing these programs and initiatives. A lack of balance in faculty's work expectations and incentives awarded for a given support initiative is likely to strain the working relationship between IT administrators and department leadership.

Similarly, grassroots projects initiated by faculty can struggle to broaden their participation beyond the parent research groups and departments. University leadership may emphasize the need for compliance and a demonstrable advantage in administrative ROI metrics (enrollment, research grants and awards, alumni giving) relative to development and support costs. The usefulness of up-and-coming EdTech tools needs to be translated into these metrics in order to build administrative support and evangelize these tools on campus.

Bridge interpretation and intervention. IT personnel often need to maintain the scope of their outreach and support initiatives to match their limited human and technical resources. Ill-defined or stopgap processes of support, however, run the risk of eroding faculty's participation and trust. An important pathway to alleviate this issue is to rigorously couple interpretation, intervention and policy iteration based on faculty's feedback. This is an exercise in (a) consolidating datasets and analyses often residing across teams within the larger IT organization, and (b) building small automations for the LMS to improve imports of legacy content, better organize course materials, improve accessibility, suggest assessment prompts, and accelerate grading. This exercise presents promising opportunities for the use of generative AI aboard the learning management system. In addition to thoroughly surveying faculty needs and technology use contexts we examined earlier (DOU, DelMo), their responsiveness to LMS automations, stagewise software rollouts, and support initiatives (instructional design cohorts, professional development programs, on-demand digital skills coursework, discussion groups) should be evaluated in a rigorously controlled fashion. The learning management system has to play a crucial role here, as it is a convenient medium to reach a representative sample of faculty (many times larger than a typical needs survey) for technology evangelism campaigns and pilot-tests of novel automations.

5.6 Limitations and Future Work

Study III contends that explanations, a transparency cue to describe stakeholder intent for recommendation users, can improve editorial trust, increase delegation, and facilitate reflection. The study examines three broad types of explanations (guarantees, social proof, content tags), and seeks to investigate which stakeholders and editorial trust contexts respond to these explanations, and why. Evidence of the unique impact of explanations on different trust contexts can empower designers to create recommendation software reusable across departments and tailor it to specific pedagogies.

Study III hypothesizes that explanations would result in overall trust gains, and social proof and content-based explanations might result in larger gains relative to simple editor guarantees. It further hypothesizes that any potential trust gains from explanations will be significantly different for low and high DOU courses (Study I), and for conservative and egalitarian trust intentions (Study II).

A subset of our survey sample is active-enrollment students and active-service faculty and teaching assistants from Summer I and Summer II terms at Virginia Tech. These courses are often small in size, their content is fast-tracked to 6 weeks, and a majority of them are undergraduate courses. Virginia Tech is a research-charter university, and summer courses are often taught by graduate students and adjunct faculty. Faculty voices are naturally underrepresented in a university-wide survey at the course-level, relative to teaching assistants and students. All of these factors can potentially affect the distribution of trust perceptions expressed by our study stakeholders, and limit the generalizability of our conclusions to peer institutions. In a future study, we plan to incorporate additional data from the next phase of our investigation in the fall of 2023, to assess the role of key variables (technology self-efficacy, peer influence, age, work experience) which potentially moderate the relationships between delegation and trust, and to conduct an online evaluation of a prototype video RS on Canvas LMS. We also intend to run controlled regression and path analyses in a full-stack study to further validate the effects of trust relationships on adoption *intentions* and *behaviors* associated with contemporary generative AIs in higher education.

Recent work on trustworthy human-AI collaboration [13] notes that untrustworthy AI coaches can successfully deceive chess players and convince them into taking misleading advice. This points to limitations of traditional XAI affordances, and the need for editorial control, resisting AI-complacency, and awareness, approval and monitoring of long-term algorithm agency.

5.7 Chapter Summary

In this section, we evaluate how faculty's editorial trust preferences translate to design implications for an educational recommender systems. Our study participants rank authorship cues (**AC**) as superior to item rationales (**IR**) and algorithm attributes (**AA**) in their potential to inspire trust and a change in editorial preferences. We also find that instructional support personnel lean towards high process disclosure (across explanation types) for medium Canvas depth-of-use as a nudge towards higher utilization of the Canvas recommender system. They also prefer frequent technology evangelism content recommendations and RS algorithm justifications for role-preserving faculty. We conclude the chapter with guidelines for UX researchers on the design of trustworthy educational recommender systems, such as dealing with trust deficits, highlighting authorship of recommended content, and ensuring the usefulness and safety of algorithmic content decisions.

Chapter 6

Conclusions

6.1 Revisiting the Studies

In the introduction to this dissertation, we outline three primary research questions. The first research question (RQI) investigates the frequent use-contexts of a learning management system, the second (RQII) examines frequent modes of editorial trust for recommendation, and the third (RQIII) evaluates design of transparency cues to facilitate editorial trust in recommender systems for teaching, learning and support. We review all salient findings from our three studies as follows:

(Study I) Needs for scale, ubiquitous access, and interoperability drive LMS adoption. Faculty's adoption of learning management systems is influenced by their perceived system efficacy at scale (ability to efficiently reach, teach and evaluate students in large classrooms, cognitive burden of transition), interoperability (ability to import and share legacy content across courses and apps), and ubiquitous access (supporting online, asynchronous, mobile learning use-cases).
See More: Section 3.5.1

(Study I) Institutional support comes in many forms. Faculty members benefit from a large array of services provided by instructional support personnel. The work roles

within institutional support include instructional designers, solution architects, accessibility specialists, content creators, project managers, and leadership, among many other. The work activities of institutional support can range from course design and retooling (design cohorts, virtual course development, learning environments) to faculty development (networked learning), technology rollouts and evangelism (campaigns, knowledgebases, licensing, grants), and software administration and support (access management, consultations, troubleshooting and technical support).
See More: Sections 1.1.1

(Study II, Study III) Instructors' preferences for editorial authority exist on a spectrum. Faculty's preferred allocation of editorial tasks for an educational recommender system range from "role-preserving" (conservative) to "role-sharing" (egalitarian). "Role-preserving" faculty tend to assign article seed, veto, refresh, and respond tasks for faculty or teaching assistants, and voting and commenting tasks to students and the RS algorithm. They cite instructor prerogatives, authorship burdens, and perceived student biases among their rationales for this allocation. "Role-sharing" faculty, in contrast, allocate all aforementioned tasks to students and the RS algorithm as necessary, citing student feedback as a vehicle for engagement, collaborative work and learning.
See More: Sections 5.4.1 and 4.4.1

(Study II, Study III) Course needs can vary widely by context. Faculty's allocation of editorial roles in student-facing recommender systems can vary considerably course type and size. Graduate, non-STEM courses with tenured faculty often assign students more autonomy as authors, allowing them to source and rate content. Large undergraduate courses with multiple sections and instructional faculty prefer limited student input, likely as active viewers who can rate and comment on course content, or as viewers who can only leave a comment on content, in order to limit the potential for spam or misinformation, and maintain course focus to meet curricular and job market expectations.

See More: Section 4.4.1

(Study III) Trust perceptions contain asymmetries. Trust beliefs (of competence, benevolence, and integrity) held by faculty, TAs, and students towards one another, contain key asymmetries. For instance, students are likely to make more optimistic assessments of staff benevolence relative to staff's self-assessments on the matter. Similarly, faculty rank student competence and benevolence significantly lower than students rank their own competence and benevolence. These trust relationships are pivotal to understanding faculty's preferred modes of RS editorial task allocation.

See More: Section 5.4.2

(Study III) Editorial authority-based explanations tend to perform better than algorithm primers and item attributes. Authorship cues typically exceed item rationales and algorithm attributes in trust, persuasiveness, and utility. They are also ranked the most likely to nudge stakeholders' overall delegation mode (from conservative to egalitarian). Trust in all three affordance groups is also consistently correlated with beliefs of student and algorithm competence, benevolence, and integrity. Faculty members tend to rank all affordance groups as less trustworthy relative to students, but this disagreement is most pronounced for authorship cues, followed by algorithm attributes and item rationales. Faculty tend to scrutinize authorship cues to ensure attribution to the primary instructor, as opposed to department colleagues or IT personnel, in curating and maintaining the recommended course materials.

See More: Section 5.4.3

(Study III) Institutional support processes need better relationship tracking, controlled technology testing, and policy transparency. The efficacy of institutional support is undermined by many factors, such as staffing constraints, poor accounting of effective strategies, balkanization of expertise, and open-loop or stopgap processes. To address these challenges, support personnel should take several steps. One, they should examine faculty engagement across all departmental communication channels (LMS announcements, departmental mailing lists, social media, websites, knowledgebases, consultations). Two, they should revise needs analysis surveys to include questions on editorial intent and domain-level benchmark tasks. Three, they should conduct controlled evaluations of faculty's reception to novel LMS automations and professional development initiatives. The findings of these evaluations should inform IT leadership on ways to both personalize and iterate faculty outreach and support policies. Learning management systems can play a promising dual role as platforms for disseminating and testing of new content, supported by a content recommender system for personalization and collaboration.

See More: Section 5.5.2

6.2 Integrating the Findings

Across its three studies, this dissertation project examines three crucial components of industrialized, efficient, and closed-loop institutional support, visualized in figure 1.1. It explores how institutional support personnel can (1) make sense of faculty's use of the learning management system and their needs, priorities, and constraints (Study I), (2) understand the initial trust relationships and delegation of tasks between faculty, TAs, students, and a LMS-hosted recommendation system (Study II), and (3) provide guidelines on selecting the recommendation system use-case, content, audiences and transparency to improve its trust-

Table 6.1: A sociotechnical decision matrix for institutional support which draws on the core design vision of this dissertation (figure 1.1). It brings together key institutional support work activities, faculty's LMS use contexts (Study I), faculty's editorial intent (Study II), and policy preferences of support personnel at Virginia Tech (Study III).

Evidence Points	**DelMo: Role-preserving**	**DelMo: Role-sharing**
DOU: Low, Medium	Strategic Focus: Minimum best practices Faculty Needs: Persuasion, digital fluency Solutions and Strategies: Automations, templates, networked learning, professional development, consultations	Strategic Focus: Minimum best practices Consensus-building Faculty Needs: Persuasion, digital fluency, student autonomy Solutions and Strategies: Automations, templates, teaching assistance
DOU: High	Strategic Focus: Consensus-building Design consistency Faculty Needs: Justification, transparency Solutions and Strategies: Professional development, peer mentoring, consultations	Strategic Focus: New tech evangelism Design consistency Faculty Needs: Justification, transparency, student autonomy Solutions and Strategies:: Automations, demos, teaching assistance

worthiness and utility as a vehicle for instructional support (Study III). It also showcases the effectiveness of full-stack strategies which take advantage of platforms already available to faculty on campus, as opposed to small-scale proofs-of-concepts which typically do not scale up and get abandoned in favor of third-party apps.

In this section, we review our analyses of policy preference survey with Virginia Tech TLOS staff (Study III) to examine how the overall strategic direction of institutional support can be informed with faculty's LMS use (DOU, Study I) and editorial intent (DelMo, Study II).

This results in a sociotechnical support matrix (see table 6.1) which describes the overall strategic focus, faculty needs, and solution approaches for all permutations of DOU and DelMo. We learn that very broadly, TLOS staff perceive faculty's challenges in technology adoption to be the lack of opportunity ('I don't have the time for it'), knowledge ('I don't know how'), confidence ('I don't believe I can get this done'), persuasion ('I don't care'), and incentive to change their habits and preferences ('I have my way of doing things'). They perceive these challenges to be among the fundamental influencing factors behind their LMS use and editorial intent. In helping them alleviate these challenges, TLOS staff push for minimum best practices and technology literacy in faculty cohorts with low DOU and low knowledge of the learning management system. In medium and high DOU cohorts, they advocate for building faculty's consensus across the cohort, and encouraging thorough use of native LMS apps beyond housekeeping tasks to meet domain-specific teaching objectives. They also recommend a focus on networked learning and consultations for role-preserving (conservative) faculty, because they are more likely to respond to institutional efforts and less likely to delegate technology discovery and management tasks to course staff, relative to role-sharing (egalitarian) faculty.

6.3 Looking Beyond

Our work aims to reveal the structural barriers to trust and process efficacy in higher education. The rise of generative AI is one crucial case study in the process effectiveness of instructional support, and a stress test for the on-campus support infrastructure. The past year has seen new generative AI tools emerge at a breathtaking pace. ChatGPT Edu [130] is a multimodal, multilingual generative AI model capable of document summarization, data analysis, and web browsing. Khanmigo [93] can generate lesson plans, end-of-lesson assessments with instructions, and create concise chunks, hooks and real-life contexts for these lessons. Harmonize AI [68] can coach students on the rubric of a given assignment, and summarize student feedback for the instructor. These up-and-coming AI tools bring the promise of unprecedented productivity gains for students, faculty and IT administrators. However, the efficacy of these tools with faculty's goals of enabling sensemaking, productivity, groupwork and assessment, is far from clear. IT divisions at major universities have commissioned AI working groups, published institutional grants, and produced compendia of self-learning resources on AI. Our research advocates for a robust, controlled UX evaluation of these tools, beginning with an assessment of benchmark tasks for an instructor's area of expertise, learning objectives, and curricular expectations. For IT organizations like TLOS, we recommend the use of usability testing artifacts and strategies (claims, work-environment models, storyboards, wireframes, rapid testing, focus groups, think-aloud sessions, heuristic evaluation) in working groups focused on AI tools. We also encourage the use of faculty's LMS platform contexts and editorial intent in order to establish a nuanced portrait of faculty needs, degree of adoption (innovators and early-adopters vs. non-adopters), and the outreach media best suited for those needs (knowledgebases, wikis, working groups, mailing lists, LMS announcements and demos, networked learning, social media).

Our work also has broader implications for HCI and CSCW research within the purview of what we identify as "institution-computer interaction". We lay the groundwork for a larger interdisciplinary conversation on work domains with a multitude of stakeholders, arranged in (1) groups with delegation and coopetition, and the presence of (2) editorial values, and (3) knowledge bounds and prerogatives. It is imperative that algorithmic decision-making in domains such as higher education, religio-spiritual storytelling, and social news be aware of the pre-existing trust perceptions, arrangements of editorial labor, and oft-competing values of domain stakeholders when their spheres of influence do not fully overlap. ICTs, especially social media, recommender systems and AI tools, can bring tremendous efficiency and democratization of information to these stakeholders. However, these information systems also create, modify, and destroy trust relationships. Tensions between academic and administrative spheres in higher education [165] and between generational strata in spiritual communities in the global south [97][98] are fundamentally linked to competing visions of open access to information and algorithmic autonomy. We hope to investigate how these tensions can be minimized, and stakeholder value-alignment enhanced, in policies for human-autonomy teaming and institutional effectiveness. We also hope to draw on influential literature in I/O psychology (organizational behavior, team attitudes, future of work, diversity) and sociocultural HCI (transnational HCI, migrant studies, postcolonial HCI, religio-spiritual communities) in describing an actionable empirical research program for HCI/CSCW research situated within institutions.

Bibliography

[1] Koofers: Professor ratings, practice exams and flash cards, jobs and internships. https://web.archive.org/web/20161217025635/https://koofers.com/. Accessed: 2024-08-28.

[2] Rate my professors. http://ratemyprofessors.com. Accessed: 2024-08-30.

[3] Solmaz Abdi, Hassan Khosravi, Shazia Sadiq, and Dragan Gasevic. Complementing educational recommender systems with open learner models. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 360–365, 2020.

[4] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1):127–158, 2020.

[5] Tella Adeyinka. Reliability and factor analysis of a blackboard course management system success: A scale development and validation in an educational context. *Journal of Information Technology Education: Research*, 10:55–80, 2011.

[6] Tella Adeyinka and Stephen Mutula. A proposed model for evaluating the success of webct course content management system. *Computers in Human Behavior*, 26(6): 1795–1805, 2010.

[7] Mohammed Ali, Trevor Wood-Harper, and Ronald Ramlogan. A framework strategy to overcome trust issues on cloud computing adoption in higher education. In *Modern*

principles, practices, and algorithms for cloud security, pages 162–183. IGI Global, 2020.

[8] Helena Alves, Emerson Wagner Mainardes, and Mário Raposo. A relationship approach to higher education institution stakeholder management. *Tertiary Education and Management*, 16:159–181, 2010.

[9] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawlat, Gautam Das, and Cong Yu. Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment*, 2(1):754–765, 2009.

[10] Kimberly E Arnold and Matthew D Pistilli. Course signals at purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 267–270. ACM, 2012.

[11] Samuel Aryee, Pawan S Budhwar, and Zhen Xiong Chen. Trust as a mediator of the relationship between organizational justice and work outcomes: Test of a social exchange model. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 23(3):267–285, 2002.

[12] Lujean Baab. Virginia Tech Rethinks Instructional Design and Faculty Development Support, 2014. https://wcet.wiche.edu/frontiers/2014/04/02/virginia-tech-rethinks-faculty-development/ [Accessed: Nov 29, 2023].

[13] Nikola Banovic, Zhuoran Yang, Aditya Ramesh, and Alice Liu. Being trustworthy is not enough: How untrustworthy artificial intelligence (ai) can deceive the end-users and gain their trust. 7(CSCW1), apr 2023. doi: 10.1145/3579460. URL https://doi.org/10.1145/3579460.

[14] AW Tony Bates and Albert Sangra. *Managing technology in higher education: Strategies for transforming teaching and learning.* John Wiley & Sons, 2011.

[15] Izak Benbasat and Weiquan Wang. Trust in and adoption of online recommendation agents. *Journal of the association for information systems*, 6(3):4, 2005.

[16] Brandford Bervell and Irfan Naufal Umar. Blended learning or face-to-face? does tutor anxiety prevent the adoption of learning management systems for distance education in ghana? *Open Learning: The Journal of Open, Distance and e-Learning*, 35(2): 159–177, 2020.

[17] Robert Birnbaum and Paul J Edelson. How colleges work: The cybernetics of academic organization and leadership, 1989.

[18] Erik W Black, Kara Dawson, and Jason Priem. Data for free: Using lms activity logs to measure community in online courses. *The Internet and Higher Education*, 11(2): 65–70, 2008.

[19] M Aaron Bond and Samantha J Blevins. Using faculty professional development to foster organizational change: A social learning framework. *TechTrends: Linking Research and Practice to Improve Learning*, 64(2):229–237, 2020.

[20] Sviatoslav Braynov and Tuomas Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, 2002.

[21] Christina Breuer, Joachim Hüffmeier, and Guido Hertel. Does trust matter more in virtual teams? a meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators. *Journal of Applied Psychology*, 101(8):1151, 2016.

[22] Ángel Alexander Cabrera, Adam Perer, and Jason I. Hong. Improving human-ai collaboration with descriptions of ai behavior. *Proc. ACM Hum.-Comput. Inter-*

act., 7(CSCW1), apr 2023. doi: 10.1145/3579612. URL https://doi.org/10.1145/3579612.

[23] Kim S Cameron. Institutional effectiveness in higher education: An introduction. *The Review of Higher Education*, 9(1):1–4, 1985.

[24] Iván Cantador and Pablo Castells. Group recommender systems: new perspectives in the social web. In *Recommender systems for the social web*, pages 139–157. Springer, 2012.

[25] Sergio W Carvalho and Márcio de Oliveira Mota. The role of trust in creating value and student loyalty in relational exchanges between higher education institutions and their students. *Journal of marketing for higher education*, 20(1):145–165, 2010.

[26] María José Casany, Marc Alier, Enric Mayol, Jordi Piguillem, Nikolas Galanis, Francisco J García-Peñalvo, Miguel Ángel Conde, et al. Moodbile: A framework to integrate m-learning applications with the lms. *Journal of Research and Practice in Information Technology*, 44(2):129, 2012.

[27] María José Casany Guerrero, Marc Alier Forment, Nikolaos Galanis, Enric Mayol Sarroca, and Jordi Piguillem Poch. Analyzing moodle/lms logs to measure mobile access. In *UBICOMM 2012: The Sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pages 35–40, 2012.

[28] Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. Supporting high-uncertainty decisions through ai and logic-style explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 251–263, 2023.

[29] Shruthi Chari, Oshani Seneviratne, Daniel M Gruen, Morgan A Foreman, Amar K

Das, and Deborah L McGuinness. Explanation ontology: A model of explanations for user-centered ai. In *International Semantic Web Conference*, pages 228–243. Springer, 2020.

[30] Weiqin Chen, Norun C Sanderson, Anna Nichshyk, Way Kiat Bong, and Siri Kessel. Usability of learning management systems for instructors–the case of canvas. In *International Conference on Human-Computer Interaction*, pages 210–223. Springer, 2021.

[31] Hamish Coates, Richard James, and Gabrielle Baldwin. A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary education and management*, 11:19–36, 2005.

[32] Mihaela Cocea and Stephan Weibelzahl. Can log files analysis estimate learners' level of motivation? In *LWA*. University of Hildesheim, Institute of Computer Science, 2006.

[33] Mihaela Cocea and Stephan Weibelzahl. Cross-system validation of engagement prediction from log files. In *European Conference on Technology Enhanced Learning*, pages 14–25. Springer, 2007.

[34] Francesco Colace, Massimo De Santo, Marco Lombardi, Rosalba Mosca, and Domenico Santaniello. A multilayer approach for recommending contextual learning paths. *Journal of Internet Services and Information Security (JISIS)*, 10(2):91–102, 2020.

[35] Shane Dawson. 'seeing'the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5):736–752, 2010.

[36] Alvise De Biasio, Andrea Montagna, Fabio Aiolli, and Nicolò Navarin. A systematic review of value-aware recommender systems. *Expert Systems with Applications*, page 120131, 2023.

[37] Carlo De Medio, Carla Limongelli, Filippo Sciarrone, and Marco Temperini. Moodlerec: A recommendation system for creating courses using the moodle e-learning platform. *Computers in Human Behavior*, 104:106168, 2020.

[38] William H DeLone and Ephraim R McLean. Information systems success: The quest for the dependent variable. *Information systems research*, 3(1):60–95, 1992.

[39] William H Delone and Ephraim R McLean. The delone and mclean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4):9–30, 2003.

[40] Sofia Balula Dias and José Alves Diniz. Towards an enhanced learning management system for blended learning in higher education incorporating distinct learners' profiles. *Journal of Educational Technology & Society*, 17(1):307–319, 2014.

[41] Hendrik Drachsler, Katrien Verbert, Olga C Santos, and Nikos Manouselis. Panorama of recommender systems to support learning. *Recommender systems handbook*, pages 421–451, 2015.

[42] Bob Edmison, Stephen H Edwards, Lujean Babb, Margaret Ellis, Chris Mayfield, Youna Jung, and Marthe Honts. Toward a new state-level framework for sharing computer science content. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pages 59–65, 2023.

[43] edutechnica. 6th Annual LMS Data Update, 2018. URL https://edutechnica.com/2018/10/06/6th-annual-lms-data-update/.

[44] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces*, pages 211–223, 2018.

[45] Asmaa Elbadrawy, R Scott Studham, and George Karypis. Collaborative multi-regression models for predicting students' performance in course activities. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 103–107. ACM, 2015.

[46] Asmaa Elbadrawy, Huzefa Rangwala, Aditya Johri, and George Karypis. SDM 2017 Tutorials: Opportunities, challenges and methods for higher education data mining, 2017. URL https://www-users.cs.umn.edu/~elbad004/sdm2017_tutorial/higher_edm_SDM_2017.pdf.

[47] Mica R Endsley. From here to autonomy: lessons learned from human–automation research. *Human factors*, 59(1):5–27, 2017.

[48] Paloma Escamilla-Fajardo, Mario Alguacil, and Samuel López-Carril. Incorporating tiktok in higher education: Pedagogical perspectives from a corporal expression sport sciences course. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 28:100302, 2021.

[49] Jo Ese. *Defending the university?: Academics' reactions to managerialism in Norwegian higher education*. PhD thesis, Karlstads universitet, 2019.

[50] Mingming Fan, Xianyou Yang, TszTung Yu, Q Vera Liao, and Jian Zhao. Human-ai collaboration for ux evaluation: Effects of explanation and synchronization. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–32, 2022.

[51] Nafsaniath Fathema, David Shannon, and Margaret Ross. Expanding the technology acceptance model (tam) to examine faculty use of learning management systems (lmss) in higher education institutions. *Journal of Online Learning & Teaching*, 11(2), 2015.

[52] Gianni Fenu, Mirko Marras, and Massimiliano Meles. A learning analytics tool for

usability assessment in moodle environments. *Journal of e-Learning and Knowledge Society*, 13(3), 2017.

[53] Rebecca Ferguson. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317, 2012.

[54] Antonio Jesús Fernández-García, Roberto Rodríguez-Echeverría, Juan Carlos Preciado, José María Conejero Manzano, and Fernando Sánchez-Figueroa. Creating a recommender system to support higher education students in the subject enrollment decision. *IEEE Access*, 8:189069–189088, 2020. doi: 10.1109/ACCESS.2020.3031572.

[55] Cynthia J Finelli, Molly Ott, Amy C Gottfried, Chad Hershock, Christopher O'neal, and Matthew Kaplan. Utilizing instructional consultations to enhance the teaching performance of engineering faculty. *Journal of Engineering Education*, 97(4):397–411, 2008.

[56] Martin Fishbein and Icek Ajzen. Belief, attitude, intention, and behavior: An introduction to theory and research. 1977.

[57] Barbara Francioni, Ilaria Curina, Charles Dennis, Savvas Papagiannidis, Eleftherios Alamanos, Michael Bourlakis, and Sabrina M Hegner. Does trust play a role when it comes to donations? a comparison of italian and us higher education institutions. *Higher Education*, 82:85–105, 2021.

[58] Gerhard Friedrich and Markus Zanker. A taxonomy for generating explanations in recommender systems. *AI Magazine*, 32(3):90–98, 2011.

[59] Andreas Fügener, Jörn Grahl, Alok Gupta, and Wolfgang Ketter. Cognitive challenges in human-ai collaboration: Investigating the path towards productive delegation. *Forthcoming, Information Systems Research*, 2019.

[60] Anja Garone, Bram Pynoo, Jo Tondeur, Celine Cocquyt, Silke Vanslambrouck, Bram Bruggeman, and Katrien Struyven. Clustering university teaching staff through utaut: Implications for the acceptance of a new learning management system. *British Journal of Educational Technology*, 50(5):2466–2483, 2019.

[61] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, and Karim Seada. Enhancing group recommendation by incorporating social relationship interactions. In *Proceedings of the 16th ACM international conference on Supporting group work*, pages 97–106, 2010.

[62] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382, 2014.

[63] David Gefen. Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 33(3):38–53, 2002.

[64] Jennifer Ann Golbeck. *Computing and applying trust in web-based social networks.* University of Maryland, College Park, 2005.

[65] Mary Sarah-Jane Gregory and Jason Michael Lodge. Academic workload: the silent barrier to the implementation of technology-enhanced learning strategies in higher education. *Distance education*, 36(2):210–230, 2015.

[66] Laurence Habib and Monica Johannesen. Perspectives on academic staff involvement in the acquisition and implementation of educational technologies. *Teaching in Higher Education*, 19(5):484–496, 2014.

[67] Anina Hanimann, Lea Hellmeuller, and Damian Trilling. Believing in credibility measures: Reviewing credibility measures in media research from 1951 to 2018. *International Journal of Communication*, 17:214–235, 2023.

[68] Harmonize Inc. Harmonize + AI for Raising Outcomes in Online Classes, 2024. URL https://harmonizelearning.com/product-page-chatgpt-in-harmonize/.

[69] Rex Hartson and Pardha S Pyla. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier, 2012.

[70] Lee Harvey and James Williams. Fifteen years of quality in higher education, 2010.

[71] Taha Hasan, Naveed Arshad, Erik Dahlquist, and Scott McCrickard. Collaborative filtering for household load prediction given contextual information. In *Proceedings of the 2017 SIAM Workshop on Machine Learning for Recommender Systems (MLRec)*, 2017.

[72] Taha Hassan. Bi-level characterization of manual setup residential non-intrusive demand disaggregation using enhanced differential evolution.

[73] Taha Hassan. On bias in social reviews of university courses. In *Companion Publication of the 10th ACM Conference on Web Science*, pages 11–14. ACM, 2019.

[74] Taha Hassan. Trust and trustworthiness in social recommender systems. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 529–532, 2019.

[75] Taha Hassan and D. Scott McCrickard. Trust and trustworthiness in social recommender systems. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion)*. ACM, 2019.

[76] Taha Hassan, Bob Edmison, Larry Cox, Matthew Louvet, and Daron Williams. Exploring the context of course rankings on online academic forums. In *2019 IEEE/ACM*

International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 553–556. IEEE, 2019.

[77] Taha Hassan, Bob Edmison, Larry Cox II, Matt Louvet, Daron Williams, and D. Scott McCrickard. Depth of use: an empirical framework to help faculty gauge the relative impact of learning management system tools. In *Proceedings of the 25th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20)*. ACM, 2020.

[78] Taha Hassan, Bob Edmison, Timothy Stelter, and D Scott McCrickard. Learning to trust: Understanding editorial authority and trust in recommender systems for education. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 24–32, 2021.

[79] Taha Hassan, Bob Edmison, Daron Williams, Larry Cox, Matthew Louvet, Bart Knijnenburg, and Donald Scott McCrickard. Simplify, consolidate, intervene: facilitating institutional support with mental models of learning management system use. In *Proceedings of the 27th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'24)*. ACM, 2024.

[80] Jiazhen He, James Bailey, Benjamin IP Rubinstein, and Rui Zhang. Identifying at-risk students in massive open online courses. In *AAAI*, pages 1749–1755, 2015.

[81] Natali Helberger, Max van Drunen, Judith Moeller, Sanne Vrijenhoek, and Sarah Eskens. Towards a normative perspective on journalistic ai: Embracing the messy reality of normative ideals, 2022.

[82] Patrick Hemmer, Monika Westphal, Max Schemmer, Sebastian Vetter, Michael Vössing, and Gerhard Satzger. Human-ai collaboration: The effect of ai delegation on

human task performance and task satisfaction. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 453–463, 2023.

[83] Curtis R Henrie, Robert Bodily, Kristine C Manwaring, and Charles R Graham. Exploring intensive longitudinal measures of student engagement in blended learning. *The International Review of Research in Open and Distributed Learning*, 16(3), 2015.

[84] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3):407–434, 2015.

[85] Kenneth Holstein and Vincent Aleven. Designing for human–ai complementarity in k-12 education. *AI Magazine*, 43(2):239–248, 2022.

[86] INSTITUTION NAME OMITTED. Upcoming Offerings, INSTITUTION NAME OMITTED, 2019. URL https://URL.OMITTED/.

[87] Osama Isaac, Adnan Aldholay, Zaini Abdullah, and Thurasamy Ramayah. Online learning usage within yemeni higher education: The role of compatibility and task-technology fit as mediating variables in the is success model. *Computers & Education*, 136:113–129, 2019.

[88] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, 4(1):53–71, 2000.

[89] Kirsten A Johnson and Burton St. John III. Transparency in the news: The impact of self-disclosure and process disclosure on the perceived credibility of the journalist, the story, and the organization. *Journalism Studies*, 22(7):953–970, 2021.

[90] James G Jones. Issues and concerns of directors of postsecondary distance learning

programs regarding online methods and technologies. *American Journal of Distance Education*, 22(1):46–56, 2008.

[91] Asra Khalid, Karsten Lundqvist, and Anne Yates. Recommender systems for moocs: A systematic literature survey (january 1, 2012–july 12, 2019). *International Review of Research in Open and Distributed Learning*, 21(4):255–291, 2020.

[92] Qasim Khalid, Naveed Arshad, Nasir Khan, Taha Hassan, Fahad Javed, and Jahangir Ikram. Home energy management system using nilm and low-cost han. *Journal of electronic science and technology*, 12(1):20–25, 2014.

[93] Khan Academy, Inc. Khanmigo: Free AI-powered Teacher Assistant by Khan Academy, 2024. URL https://www.khanmigo.ai/teachers.

[94] Mohamed Koutheaïr Khribi, Mohamed Jemni, and Olfa Nasraoui. Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, pages 241–245. IEEE, 2008.

[95] Bart P Knijnenburg and Martijn C Willemsen. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*, pages 309–352. Springer, 2015.

[96] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, 22(4):441–504, 2012.

[97] Lindah Kotut, Neelma Bhatti, Morva Saaty, Derek Haqq, Timothy L Stelter, and D Scott McCrickard. Clash of times: Respectful technology space for integrating community stories in intangible exhibits. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[98] Lindah Kotut, Neelma Bhatti, Taha Hassan, Derek Haqq, and Morva Saaty. Griot-style methodology: Longitudinal study of navigating design with unwritten stories. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2024.

[99] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[100] John Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992.

[101] Brian Lubars and Chenhao Tan. Ask not what ai can do, but what ai should do: Towards a framework of task delegability. *Advances in Neural Information Processing Systems*, 32, 2019.

[102] Leah P Macfadyen and Shane Dawson. Mining lms data to develop an "early warning system" for educators: A proof of concept. *Computers & education*, 54(2):588–599, 2010.

[103] Cathryn A Manduca, Ellen R Iverson, Michael Luxenberg, R Heather Macdonald, David A McConnell, David W Mogk, and Barbara J Tewksbury. Improving undergraduate stem education: The efficacy of discipline-based professional development. *Science Advances*, 3(2):e1600193, 2017.

[104] Nikos Manouselis, Hendrik Drachsler, Riina Vuorikari, Hans Hummel, and Rob Koper. Recommender systems in technology enhanced learning. In *Recommender systems handbook*, pages 387–415. Springer, 2011.

[105] Jin Mao, Baiyun Chen, and Juhong Christie Liu. Generative artificial intelligence in education and its implications for assessment. *TechTrends*, 68(1):58–66, 2024.

[106] Roxana Marachi and Lawrence Quill. The case of canvas: Longitudinal datafication through learning management systems. *Teaching in Higher Education*, 25(4):418–434, 2020.

[107] Maria Rosala. How to Analyze Qualitative Data from UX Research: Thematic Analysis, 2024. URL https://www.nngroup.com/articles/thematic-analysis/.

[108] Florence Martin, Drew Polly, Annika Jokiaho, and Birgit May. Global standards for enhancing quality in online learning. *Quarterly Review of Distance Education*, 18(2): 1–102, 2017.

[109] Florence Martin, Ting Sun, and Carl D Westine. A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & education*, 159:104009, 2020.

[110] Ludivine Martin, David Roldán Martínez, Olga Revilla, Maria J Aguilar, Olga C Santos, and Jesús G Boticario. Usability in e-learning platforms: heuristics comparison between moodle, sakai and dotlrn. In *The 7th Europian Conference on e-Learning, Agia Napa, Cyprus*. Citeseer, 2008.

[111] Susana Masapanta-Carrión and J Ángel Velázquez-Iturbide. A systematic review of the use of bloom's taxonomy in computer science education. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 441–446. ACM, 2018.

[112] Paolo Massa and Paolo Avesani. Trust-aware collaborative filtering for recommender systems. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 492–508. Springer, 2004.

[113] Paolo Massa and Paolo Avesani. Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 17–24, 2007.

[114] Margaret Mattes. The private side of public higher education. 2017.

[115] Riccardo Mazza and Christian Milani. Gismo: a graphical interactive student monitoring tool for course management systems. In *International Conference on Technology Enhanced Learning, Milan*, pages 1–8, 2004.

[116] Tanya J McGill and Jane E Klobas. A task–technology fit view of learning management system impact. *Computers & Education*, 52(2):496–508, 2009.

[117] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3):334–359, 2002.

[118] Laura-Ann Migliore. Leadership, governance, and perceptions of trust in the higher education industry. *Journal of Leadership Studies*, 5(4):30–40, 2012.

[119] Vladimir Mikić, Miloš Ilić, Lazar Kopanja, and Boban Vesin. Personalisation methods in e-learning-a literature review. *Computer Applications in Engineering Education*, 30 (6):1931–1958, 2022.

[120] Susan D Moisey and Judith A Hughes. Supporting the online learner. *Theory and practice of online learning,*, pages 419–439, 2008.

[121] Cedric Bheki Mpungose and Simon Bheki Khoza. Postgraduate students' experiences on the use of moodle and canvas learning management system. *Technology, Knowledge and Learning*, 27(1):1–16, 2022.

[122] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pages 2431–2439. IEEE, 2002.

[123] VT News. Virginia Tech Chooses Canvas as Next Generation Learning Management System, 2015. https://news.vt.edu/articles/2015/06/060215-it-canvasannouncement.html/ [Accessed: Nov 29, 2023].

[124] Eric WT Ngai, JKL Poon, and YHC Chan. Empirical examination of the adoption of webct using tam. *Computers & Education*, 48(2):250–267, 2007.

[125] Thao Ngo, Johannes Kunkel, and Jürgen Ziegler. Exploring mental models for transparent and controllable recommender systems: a qualitative study. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 183–191, 2020.

[126] Samuel Nowakowski, Ivana Ognjanović, Monique Grandbastien, Jelena Jovanovic, and Ramo Šendelj. Two recommending strategies to enhance online presence in personal learning environments. *Recommender systems for technology enhanced learning: research trends and applications*, pages 227–249, 2014.

[127] Melanie T Older, Patrick E Waterson, and Chris W Clegg. A critical assessment of task allocation methods and their applicability. *Ergonomics*, 40(2):151–171, 1997.

[128] Jeroen Ooge, Shotallo Kato, and Katrien Verbert. Explaining recommendations in e-learning: Effects on adolescents' trust. In *27th International Conference on Intelligent User Interfaces*, pages 93–105, 2022.

[129] Jeroen Ooge, Leen Dereu, and Katrien Verbert. Steering recommendations and visual-

ising its impact: Effects on adolescents' trust in e-learning platforms. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 156–170, 2023.

[130] OpenAI, Inc. Introducing ChatGPT Edu, 2024. URL https://openai.com/index/introducing-chatgpt-edu/.

[131] Sevgi Ozkan and Refika Koseler. Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation. *Computers & Education*, 53(4):1285–1296, 2009.

[132] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, 64(5):904–938, 2022.

[133] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297, 2000.

[134] Prajish Prasad. Developing students' cognitive processes required for software design verification. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*, pages 284–285. ACM, 2018.

[135] Jaziar Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & education*, 147:103778, 2020.

[136] Yosef S Razin and Karen M Feigh. Converging measures and an emergent model: A meta-analysis of human-automation trust questionnaires. *arXiv preprint arXiv:2303.13799*, 2023.

[137] Bart Rienties, Natasa Brouwer, and Simon Lygo-Baker. The effects of online professional development on higher education teachers' beliefs and intentions towards learning facilitation and technology. *Teaching and teacher education*, 29:122–131, 2013.

[138] S Kay Rockwell, Jolene Schauer, Susan Fritz, and David B Marx. Incentives and obstacles influencing higher education faculty and administrators to teach via distance. *Faculty Publications: Agricultural Leadership, Education & Communication Department*, page 53, 1999.

[139] Cristóbal Romero and Sebastián Ventura. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618, 2010.

[140] Cristóbal Romero, Sebastián Ventura, and Enrique García. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51 (1):368–384, 2008.

[141] Yao Rong, Tobias Leemann, Thai-trang Nguyen, Lisa Fiedler, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. Towards human-centered explainable ai: User studies for model explanations. *arXiv preprint arXiv:2210.11584*, 2022.

[142] Guido Rößling and Andreas Kothe. Extending moodle to better support computing education. *ACM SIGCSE Bulletin*, 41(3):146–150, 2009.

[143] Guido Rößling, Myles McNally, Pierluigi Crescenzi, Atanas Radenski, Petri Ihantola, and M Gloria Sánchez-Torrubia. Adapting moodle to better support cs education. In *Proceedings of the 2010 ITiCSE working group reports*, pages 15–27. ACM, 2010.

[144] Morva Saaty, Derek Haqq, Mohammadreza Beyki, Taha Hassan, and D SCOTT McCRICKARD. Pokémon go with social distancing: Social media analysis of players'

experiences with location-based games. *Proceedings of the ACM on Human-Computer Interaction*, 6(CHI PLAY):1–22, 2022.

[145] Asher Irfan Saroia and Shang Gao. Investigating university students' intention to use mobile learning management systems in sweden. *Innovations in Education and Teaching International*, 56(5):569–580, 2019.

[146] Walesska Schlesinger, Amparo Cervera, and Carmen Pérez-Cabañero. Sticking with your university: the importance of satisfaction, trust, image, and shared values. *Studies in Higher Education*, 42(12):2178–2194, 2017.

[147] Shunichi Seko, Takashi Yagi, Manabu Motegi, and Shinyo Muto. Group recommendation using feature space representing behavioral tendency and power balance among members. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 101–108, 2011.

[148] Neil Selwyn, Michael Henderson, and Shu-Hua Chao. 'you need a system': Exploring the role of data in the administration of university students and courses. *Journal of Further and Higher Education*, 42(1):46–56, 2018.

[149] Baris Sezer and Ramazan Yilmaz. Learning management system acceptance scale (lmsas): A validity and reliability study. *Australasian Journal of Educational Technology*, 35(3), 2019.

[150] Thomas B Sheridan and Raja Parasuraman. Human-automation interaction. *Reviews of human factors and ergonomics*, 1(1):89–129, 2005.

[151] Sonali Singh and Sumeet Singh Jasial. Moderating effect of perceived trust on service quality–student satisfaction relationship: evidence from indian higher management

education institutions. *Journal of Marketing for Higher Education*, 31(2):280–304, 2021.

[152] Traci Sitzmann, Katherine Ely, Bradford S Bell, and Kristina N Bauer. The effects of technical difficulties on learning and attrition during online training. *Journal of Experimental Psychology: Applied*, 16(3):281, 2010.

[153] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. Sensecape: Enabling multi-level exploration and sensemaking with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–18, 2023.

[154] Lu Sun, Stone Tao, Junjie Hu, and Steven P. Dow. Metawriter: Exploring the potential and perils of ai writing support in scientific peer review. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1), apr 2024. doi: 10.1145/3637371. URL https://doi.org/10.1145/3637371.

[155] Mary K Tallent-Runnels, Julie A Thomas, William Y Lan, Sandi Cooper, Terence C Ahern, Shana M Shaw, and Xiaoming Liu. Teaching courses online: A review of the research. *Review of educational research*, 76(1):93–135, 2006.

[156] Nava Tintarev and Judith Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*, pages 479–510. Springer, 2011.

[157] Cathrine Edelhard Tømte, Trine Fossland, Per Olaf Aamodt, and Lise Degn. Digitalisation in higher education: mapping institutional approaches for teaching and learning. *Quality in Higher Education*, 2019.

[158] US Department of Homeland Security. STEM Designated Degree Program List,

2016. URL https://www.ice.gov/sites/default/files/documents/Document/2016/stem-list.pdf.

[159] Patricia Victor, Martine De Cock, and Chris Cornelis. Trust and recommendations. In *Recommender systems handbook*, pages 645–675. Springer, 2011.

[160] Lesley Vidovich and Jan Currie. Governance and trust in higher education. *Studies in Higher Education*, 36(1):43–56, 2011.

[161] Qian Wan, Siying Hu, Yu Zhang, Piaohong Wang, Bo Wen, and Zhicong Lu. "it felt like having a second mind": Investigating human-ai co-creativity in prewriting with large language models. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1), apr 2024. doi: 10.1145/3637361. URL https://doi.org/10.1145/3637361.

[162] Weiquan Wang and Izak Benbasat. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, 23(4):217–246, 2007.

[163] Bryan R Warnick. *Imitation and education: A philosophical inquiry into learning by example*. SUNY Press, 2009.

[164] Richard Watermeyer, Tom Crick, Cathryn Knight, and Janet Goodall. Covid-19 and digital disruption in uk universities: Afflictions and affordances of emergency online migration. *Higher education*, 81:623–641, 2021.

[165] John F Welsh and Jeff Metcalf. Faculty and administrative support for institutional effectiveness activities: a bridge across the chasm? *The Journal of Higher Education*, 74(4):445–468, 2003.

[166] Richard E West, Greg Waddoups, and Charles R Graham. Understanding the experi-

ences of instructors as they adopt a course management system. *Educational Technology Research and Development*, 55(1):1–26, 2007.

[167] Diane Wilcox, Jane Thall, and Oris Griffin. One canvas, two audiences: How faculty and students use a newly adopted learning management system. In *Society for Information Technology & Teacher Education International Conference*, pages 1163–1168. Association for the Advancement of Computing in Education (AACE), 2016.

[168] Daron Williams, Larry Cox, Margaret Ellis, Bob Edmison, Taha Hassan, M. Aaron Bond, Quinn Warnick, Virginia Clark, Daniel Yaffe, Molly Domino, and Derek Haqq. *Data-Informed Learning Design in a Computer Science Course*, pages 1–23. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-319-17727-4_176-1. URL https://doi.org/10.1007/978-3-319-17727-4_176-1.

[169] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 145–149. ACM, 2013.

[170] Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS quarterly*, pages 137–209, 2007.

[171] Olaf Zawacki-Richter, Eva Maria Bäcker, and Sebastian Vogt. Review of distance education research (2000 to 2008): Analysis of research areas, methods, and authorship patterns. *International Review of Research in Open and Distributed Learning*, 10(6): 21–50, 2009.

Appendix A

Publications

The following manuscripts are accepted (**A**), under peer review (**PR**), or planned (**P**) within the scope of Studies I through III.

(**CSCW'24**, **A**) **Taha Hassan**, Bob Edmison, Daron Williams, Larry Cox II, Matthew Louvet, Bart Knijnenburg, and D. Scott McCrickard (2024). Simplify, Consolidate, Intervene: Facilitating Institutional Support with Mental Models of Learning Management System Use. In Proceedings of the 27th ACM Conference on Computer-Supported Cooperative Work.

(**Springer LDT'22**, **A**) Daron Williams, Larry Cox II, Margaret Ellis, Bob Edmison, **Taha Hassan**, Aaron Bond, Virginia Clark, Daniel Yaffe, Molly Domino and Derek Haqq (2022). Data-Informed Learning Design in a Computer Science Course. In: Spector, M.J., Lockee, B.B., Childress, M.D. (eds) Learning, Design, and Technology. Springer. [Book Chapter]

(**UMAP'21**, **A**) **Taha Hassan**, Bob Edmison, Timothy Stelter, and D. Scott McCrickard. "Learning to Trust: Understanding Editorial Authority and Trust in Recommender Systems for Education." In Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, pp. 24-32. 2021. [Acceptance Rate: 23.3%]

(**ITiCSE'20**, **A**) **Taha Hassan**, Bob Edmison, Larry Cox, Matt Louvet, Daron Williams, and D. Scott McCrickard. "Depth of Use: An Empirical Framework to Help Faculty Gauge the Relative Impact of Learning Management System Tools." In Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education, pp. 47-53.

2020. [Acceptance Rate: 28%]

(**WWW'19 Companion**, **A**) **Taha Hassan** and D. Scott McCrickard. "Trust and trustworthiness in social recommender systems." In Companion Proceedings of The 2019 World Wide Web Conference, pp. 529-532. 2019.

(**WebSci'19 Companion, A**) **Taha Hassan**. "On bias in social reviews of university courses." In Companion Publication of the 10th ACM Conference on Web Science, pp. 11-14. 2019.

(**ASONAM'19, A**) **Taha Hassan**, Bob Edmison, Larry Cox, Matthew Louvet, and Daron Williams. "Exploring the context of course rankings on online academic forums." In 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 553-556. IEEE, 2019.

(**CSCW'24, PR**) **Taha Hassan**, D. Scott McCrickard, Bob Edmison, Kristylea Ojeda, Sally Hamouda, Michael Schulz, Alexandra Hyler, Donna Fortune, Christian Lucero, Ted Price and Maha Elouni. "The contours of prerogative: trustworthy editorial arrangements in recommendation of learning resources". Under revision for 2024 ACM Conference on Computer-Supported Cooperative Work.

Appendix B

DOU Hypothesis Tests

Tables B.1, B.2, B.3 and B.4 describe the hypothesis-tests for DOU and course metadata from Fall 2017 through Fall 2022 terms at Virginia Tech.

Table B.1: Fall 2017 and Spring 2018: Hypothesis-testing the relationship between DOU and key course attributes.

Hypothesis	**(F17)** t, F, H	**(S18)** t, F, H
H1: Undergraduate	9**, 81.1**, 75.8**	9**, 80.3**, 75.3**
H2: STEM	-4.5*, 19.8*, 22.4*	-4.5*, 20.4*, 20.8*
H3: Online	0.5, 0.3, 0	4.6*, 21.5*, 19.5*
H4: App use	32.3***, 1e3***, 1e3***	45.3*, 2e3*, 1e3*
H5: Enrollment	- , 69.1**, 974.8***	- , 73***, 969***
H6: Viewership	- , 4.4*, 3.6	- , 4.2*, 0.5
H7: GPA	- , 3.8*, 4.2	- , 3.6*, 2.8
H8: DFW	- , 5.3*, 2.5	- , 3*, 0.3
H9: #TA	- , 108.5***, 1e3***	- , 91***, 1e3***
H10: Skills	6.1*, 37.7*, 52.8**	3.7*, 13.8*, 20.9*

$\alpha = 0.05$, stat. signif. $p < \alpha \wedge$ $^*p > 1e{-}10$, $^{**}p > 1e{-}20$, $^{***}p < 1e{-}20$

Table B.2: Fall 2018, Spring 2019 and Fall 2019: Hypothesis-testing the relationship between DOU and key course attributes

Hyp.	**(F18)** t, F, H	**(S19)** t, F, H	**(F19)** t, F, H
H1	5.8*, 33.6*, 33.1*	6.3*, 40*, 39*	21.4***, 4e2***, 4e2***
H2	0.9, 0.7, 0.6	-0.8, 0.6, 0.7	6.4**, 40.6**, 45.1***
H3	2.2*, 4.7*, 5.3*	1.5, 2.4, 2.9	-1.4, 2.1, 1.8
H4	1.1, 1.1, 0.9	1.4, 2.0, 1.8	36.3***, 1e3***, 1e3***
H5	- , 9.7*, 170.9***	- , 10.2*, 181.9***	- , 95.1***, 1e3***
H6	- , 5.9*, 4.2	- , 1.5, 2.6	- , 1.5, 8.7*
H7	- , 7.8*, 10.4*	- , 1.5, 0.4	- , 0.7, 0.6
H8	- , 7*, 1.3	- , 3.5*, 2.7	- , 0.6, 0.8
H9	- , 5.9*, 76.8**	- , 3.4*, 61.7**	- , 146.5***, 1e3***
H10	-3.6*, 12.7*, 13*	-5*, 25*, 25*	6.9**, 48.1**, 73.2**

$\alpha = 0.05$, stat. signif. $p < \alpha \wedge$ $^*p > 1e{-}10$, $^{**}p > 1e{-}20$, $^{***}p < 1e{-}20$

Table B.3: Spring 2021 and Fall 2021: Hypothesis-testing the relationship between DOU and key course attributes.

Hypothesis	**(S21)** t, F, H	**(F21)** t, F, H
H1: Undergraduate	9.4**, 88.6**, 88.7**	11.5**, 131.2**, 128.6**
H2: STEM	-5.2*, 27*, 26*	-5*, 26.8*, 25.7*
H3: Online	8.6**, 74.7**, 68**	3.9*, 15.2*, 15.8*
H4: App use	3.3*, 11*, 12.5*	5.5*, 29.7*, 30.2*
H5: Enrollment	-, 89.6**, 1e3**	-, 66.6**, 1e3**
H6: Viewership	-, 1e3**, 2e3**	-, 809**, 2e3**
H7: GPA	-, 81.8**, 187**	-, 24**, 63.5**
H8: DFW	-, 31.5**, 130**	-, 12*, 98**
H9: #TA	-, 43**, 488**	-, 31**, 477**
H10: Skills	-2*, 4.2*, 4.1*	1.9*, 3.8*, 4*

*$\alpha = 0.05$, stat. signif. $p <= \alpha \wedge p > 1e{-}10$, **$p < 1e{-}10$

Table B.4: Spring 2022 and Fall 2022: Hypothesis-testing the relationship between DOU and key course attributes.

Hypothesis	**(S22)** t, F, H	**(F22)** t, F, H
H1: Undergraduate	9.3**, 87.6**, 87.6**	11.6**, 135**, 132.8**
H2: STEM	-4.1*, 17.2*, 16.8*	-6.6**, 44.5**, 42.8**
H3: Online	2*, 4*, 4.3*	5*, 25.4*, 25.9*
H4: App Use	6.9**, 47.5**, 47.4**	7.6**, 58.7**, 59**
H5: Enrollment	-, 82**, 1.1e3**	-, 87.7**, 1e3**
H6: Viewership	-, 267**, 2e3**	-, 1e3**, 2e3**
H7: GPA	-, 35.7**, 106**	-, 31.7**, 93.2**
H8: DFW	-, 12.2*, 95.4**	-, 18.2*, 1e2**
H9: #TA	-, 65.5**, 462**	-, 51**, 447**
H10: Skills	0.48, 0.23, 0.2	1.3, 1.7, 1.9

*$\alpha = 0.05$, stat. signif. $p <= \alpha \wedge p > 1e{-}10$, **$p < 1e{-}10$

Appendix C

DOU Regression Tests

Tables C.1, C.2, C.3, and C.4 present the results of OLS regression, with DOU as the dependent variable, and course attributes from Study I as independent variables. These include audiences (Undergraduate, STEM), modality (online, app use), participation (student count, viewership), logistics (TA count, NLI) and outcomes (GPA).

Table C.1: (Spring 2021) OLS Regression Results, Dep. Variable: DOU, R-squared: 0.170, Model: OLS, Adj. R-squared: 0.168, Method: Least Squares, F-statistic: 73.93, Prob (F-statistic): 1.22e-124, Log-Likelihood: -2477.3, No. Observations: 3248, AIC: 4975, Df Residuals: 3238, BIC: 5035, Df Model: 9.

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.8206	0.102	8.035	0.000	0.620	1.021
gpa	-0.0023	0.024	-0.096	0.923	-0.050	0.045
undergrad	0.1236	0.026	4.764	0.000	0.073	0.174
stem	-0.1174	0.019	-6.034	0.000	-0.156	-0.079
online	-0.1040	0.021	-4.937	0.000	-0.145	-0.063
studentcount	0.0003	0.0000925	3.356	0.001	0.000	0.000
ta	0.0099	0.003	3.449	0.001	0.004	0.016
viewership	0.0004	0.0000172	21.337	0.000	0.000	0.000
nli	-0.0258	0.020	-1.310	0.190	-0.064	0.013
app_use	-0.2196	0.042	-5.288	0.000	-0.301	-0.138

Table C.2: (Fall 2021) OLS Regression Results, Dep. Variable: DOU, R-squared: 0.153, Model: OLS, Adj. R-squared: 0.151, Method: Least Squares, F-statistic: 68.34, Prob (F-statistic): 4.44e-116, Log-Likelihood: -2629.7, No. Observations: 3405, AIC: 5279, Df Residuals: 3395, BIC: 5341, Df Model: 9.

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.7362	0.094	7.822	0.000	0.552	0.921
gpa	0.0242	0.022	1.117	0.264	-0.018	0.067
undergrad	0.1800	0.026	6.857	0.000	0.129	0.231
stem	-0.1547	0.019	-7.983	0.000	-0.193	-0.117
online	-0.0706	0.025	-2.817	0.005	-0.120	-0.021
studentcount	0.0002	0.0000744	3.159	0.002	0.0000891	0.0003
ta	0.0065	0.002	2.802	0.005	0.002	0.011
viewership	0.0004	0.0000174	20.560	0.000	0.000	0.000
nli	0.0035	0.019	0.188	0.851	-0.033	0.040
app_use	-0.0120	0.043	-0.278	0.781	-0.097	0.073

Table C.3: (Spring 2022) OLS Regression Results, Dep. Variable: DOU, R-squared: 0.085, Model: OLS, Adj. R-squared: 0.083, Method: Least Squares, F-statistic: 34.66, Prob (F-statistic): 5.32e-59, Log-Likelihood: -2674.4, No. Observations: 3363, AIC: 5369, Df Residuals: 3353, BIC: 5430, Df Model: 9.

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	1.2345	0.092	13.418	0.000	1.054	1.415
gpa	-0.0610	0.021	-2.836	0.005	-0.103	-0.019
undergrad	0.0956	0.028	3.457	0.001	0.041	0.150
stem	-0.1326	0.020	-6.610	0.000	-0.172	-0.093
online	-0.0516	0.027	-1.931	0.054	-0.104	0.001
studentcount	0.0002	9.4e-05	1.859	0.063	-9.59e-06	0.000
ta	0.0166	0.003	5.807	0.000	0.011	0.022
viewership	0.0001	8.5e-06	12.311	0.000	8.8e-05	0.000
nli	-0.0076	0.019	-0.392	0.695	-0.045	0.030
app_use	0.0453	0.045	0.999	0.318	-0.044	0.134

Table C.4: (Fall 2022) OLS Regression Results, Dep. Variable: DOU, R-squared: 0.182, Model: OLS, Adj. R-squared: 0.180, Method: Least Squares, F-statistic: 85.45, Prob (F-statistic): 7.57e-144, Log-Likelihood: -2556.0, No. Observations: 3472, AIC: 5132, Df Residuals: 3462, BIC: 5194, Df Model: 9.

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.6072	0.091	6.698	0.000	0.429	0.785
gpa	0.0636	0.021	3.090	0.002	0.023	0.104
undergrad	0.1679	0.025	6.627	0.000	0.118	0.218
stem	-0.1794	0.019	-9.624	0.000	-0.216	-0.143
online	-0.1159	0.026	-4.493	0.000	-0.166	-0.065
studentcount	0.0003	8.08e-05	3.615	0.000	0.000	0.000
ta	0.0127	0.003	4.320	0.000	0.007	0.019
viewership	0.0004	1.62e-05	22.760	0.000	0.000	0.000
nli	-0.0124	0.017	-0.711	0.477	-0.047	0.022
app_use	0.0140	0.037	0.379	0.705	-0.058	0.086

Appendix D

Survey Instruments

D.1 Pilot Study (Work Practice of Institutional Support)

1. What are "technology-enhanced learning" and "online strategies"? How many qualitatively different institutional support services exist across TLOS?
2. What are the key work roles at TLOS?
3. How would you situate your work role and responsibilities within the TLOS organization?
4. What are the frequent challenges with the overall work practice of TLOS (course retooling, tech evangelism and outreach, support allocation and management, and others)?
5. Think about the gaps in current technology at TLOS (for course redesign/retooling/certification, technology rollouts and evangelism, support allocation and management) or unmet user needs, and how the following technologies address the gaps, or meet these needs:
Canvas use analytics, course personas, recommender systems for Canvas, faculty relationship management, decision analytics. (Strongly Disagree-Strongly Agree)
6. Tell us which of the aforementioned technologies is potentially the most impactful for instructional support (course redesign, learning environments, tech evangelism, etc), and why. Feel free to talk about the human, tech or institutional challenges you think this technology will address.

D.2 Study II-III (Trust, Delegation and Transparency)

Note: Study II uses single-item trust questions with regards to each stakeholder (course instructor, teaching assistants, students) and the response values ranging from low to high (5 points). Other questions appearing in the Study II survey are noted with an asterisk (*) in the following list. All of the questions in the list appear in Study III.

D.2.1 Trust Beliefs

(Strongly Disagree-Strongly Agree)

(Belief: Competence)

1. [Course Instructors, Teaching Assistants, Students] have the ability to understand my need of reading materials for this course.

2. [Course Instructors, Teaching Assistants, Students] have good knowledge about the course domain and topics.

(Belief: Benevolence)

3. [Course Instructors, Teaching Assistants, Students] will put my interests first in suggesting reading materials for this course.

4. [Course Instructors, Teaching Assistants, Students] keep my interests in mind.

(Belief: Integrity)

5. [Course Instructors, Teaching Assistants, Students] are honest.

6. [Course Instructors, Teaching Assistants, Students] have integrity.

D.2.2 Intent to Delegate Editorial Authority

1. *In your opinion, who can *seed* readings for this Suggested Readings (refer to fig.) recommender system? Seed readings are a handful of example readings to help set up the algorithm that finds additional, similar readings.

2. *Who, if any, should have the *veto power* in recommendation process? This guarantees the power to instantly remove a recommended reading.

3. *Who, if any, should be able to *like*, *dislike*, *upvote*, or *downvote* the recommended readings? This gives them *soft power* because these likes/dislikes could be accounted for by the recommendation algorithm in moving the readings up and down the list.

4. *Who should be able to comment on, or provide feedback on a recommended reading?

5. Who should be able to refresh the recommendations?

6. Who should receive, and when applicable, respond to notifications from the recommendation system, for instance, likely vetoes, new recommended sources, admin actions, student comments, or event-based triggers?

D.2.3 Transparency Affordances: Impressions, Barriers, Impact

1. Do you think this [item rationale, authorship cue, algorithm attribute] is [effective, trustworthy, useful, persuasive]? (Strongly Disagree-Strongly Agree)

2. Why do you think this [item rationale, authorship cue, algorithm attribute] might not be trustworthy?

3. What is one change you would make to this [item rationale, authorship cue, algorithm attribute] to enhance its trustworthiness?

4. Here are all the [item rationales, authorship cues, algorithm attributes] you have examined so far. Do you agree with the following?

(a) This transparency affordance makes me reflect on how recommendation algorithms can assist in suggesting course readings.
(b) This transparency affordance makes me reflect on how instructors, teaching assistants and students can help recommend readings.
(c) This transparency affordance might make me change my mind about including instructors, teaching assistants and students in the recommendation process.
(d) This transparency affordance might make me trust the algorithm more. (Strongly Disagree-Strongly Agree)

D.2.4 Expertise

(Strongly Disagree-Strongly Agree) 1. I have a thorough understanding of, or expertise in recommendation systems.
2. I have a thorough understanding of how recommendation systems like Netflix, YouTube or TikTok suggest content to me.

D.2.5 Leadership

(Strongly Disagree-Strongly Agree) 1. I have a thorough understanding of, or expertise in recommendation systems.
2. I have a thorough understanding of how recommendation systems like Netflix, YouTube or TikTok suggest content to me. 3. I have a thorough understanding of how recommendation systems like Netflix, YouTube or TikTok suggest content to me. 4. I have a thorough understanding of how recommendation systems like Netflix, YouTube or TikTok suggest content to me.

D.2.6 Preference for Automation

*In the [short-term, long-term], how do you view the relative role of the recommendation algorithm and human feedback (instructor, TAs or students) in providing recommendations you can trust? Select one of [all human; mostly human, some algorithm; equal; mostly algorithm, some human; all algorithm].

D.3 Study III (Policy Preferences Survey)

1. Suppose we assemble a metric of faculty's overall Canvas use (low, medium or high). This metric is built using a faculty member's use of Canvas announcements, syllabus, files, assignments, quizzes, gradebook, and discussions. We can then use this metric to select and personalize content for a recommender system. For example, it can be a "Virginia Tech recommends", "TLOS recommends", or "Your department recommends"-style feed on the Canvas dashboard, with faculty as primary audiences. Or, it can be a "Recommended readings" or "Your instructor recommends"-style feed on the Canvas course homepage or the "modules" pages, with students enrolled in a class as primary audiences.
What kind of design choices (audiences, content, transparency, customizability) would you make for this recommender system if the course faculty had with different levels of Canvas use?
2. Suppose we assemble a metric of faculty's delegation of tasks to TAs, students and AI/algorithms. Our analyses tell us that two broad camps of faculty members exist on a spectrum of editorial preferences from conservative to egalitarian. Conservative or "role-preserving" faculty tend to teach large, multi-section undergraduate courses. They are hesitant in giving away editorial power to students, TAs and AI/algorithms. They are often concerned about student disengagement and misinformation. Egalitarian or "role-sharing" faculty tend to

teach smaller, graduate courses. In delegating tasks to TAs, students and algorithms, they are invested in student engagement, and tend to trust them more than ”conservative” faculty.

What kind of design choices (audiences, content, transparency, customizability) would you make for this recommender system (mentioned above) if the primary course instructor was conservative (role-preserving) vs. egalitarian (role-sharing)?

Appendix E

Interview Questions

E.1 Study I (Depth-of-Use)

The discussion questions from Study I (steps 0 and 1), listed as follows, facilitated the evolution of Canvas analytics artifacts, and derivation of the DOU taxonomy.

1. What are the pros and cons of the given Canvas data analytics artifact (figure 3.2) in helping evaluate a faculty member's use of Canvas?

2. What in your view counts as low, medium, and high use of Canvas services [announcements, quizzes, assignments, discussions, gradebook, syllabus, files]?

3. Which Canvas services will you combine pairwise? Why? Use the truth table (provided) to mark the overall use of each pair of services as 1 or 0.

E.2 Study II-III (Trust, Delegation, and Recommender Systems)

(These accompany the survey questions in Appendix D.2)

1. What are your reasons for how you chose to allocate recommendation tasks (of *seed*, *veto*, *rate*, *comment*, *refresh*, and *notify*) in the survey, to faculty, students, teaching assistants,

and the RS algorithm?

2. What are the reasons for the trust beliefs you expressed in the survey regarding faculty, students, teaching assistants, and the RS algorithm?

3. What are the reasons for your preferred level of automation for the recommender system, both in the short-term and long-term?