IEEE TRANSACTIONS ON

PROFESSIONAL COMMUNICATION

A PUBLICATION OF THE IEEE PROFESSIONAL COMMUNICATION SOCIETY

IEEE PRO COMM Professional Communication Society

DECEMBER 2021 VOLUME 64 NUMBER 4 IEPCBU (ISSN 0361-1434)

IEEE

IEEE PROFESSIONAL COMMUNICATION SOCIETY

The IEEE Professional Communication Society is an organization within the framework of IEEE that is dedicated to improving the communication of technical information. For membership and subscription information and pricing, please visit https://www.ieee.org/membership-catalog.

IEEE Transactions on Professional Communication (ISSN 1558-1500) is published quarterly by The Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Corporate Office**: 3 Park Avenue, 17th Floor, New York, NY 10016. **IEEE Operations Center**: 445 Hoes Lane, Piscataway, NJ 08854. **NJ Telephone**: +1 732 981 0060. **Price/Publication Information**: Member and nonmember subscription prices available on request. **Copyright and Reprint Permissions**: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use of patrons, provided the per-copy fee of $31.00 is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For all other copying, reprint, or republication permission, write to Copyrights and Permissions Department, IEEE Publications Administration, 445 Hoes Lane, Piscataway, NJ 08854-4141.

Digital Object Identifier 10.1109/TPC.2021.3127619

Integrative Literature Review

Elevator Pitch Assessment Model: A Systematization of Dimensions in Technology Entrepreneurship Presentations

—Alessandro Margherita and David Verrill

***Abstract*—Introduction:** *The creation of a technology venture brings the entrepreneur to interact with different stakeholders and persuade them of the quality of the business idea. In such endeavors, entrepreneurial storytelling and business pitches are crucial to attract stakeholder interest and potential commitment. We focus on longer and structured elevator pitches used by entrepreneurs seeking funds and partners for their startup, and we present an integrative framework of evaluation dimensions, specific items, and key evidence to assess a pitch.* **Research methodology:** *We conducted a systematic review of specialized literature on business venturing, entrepreneurship, and business communication, and we selected 40 research articles from which we have extracted concepts related to the quality and effectiveness of an elevator pitch. We analyze and aggregate concepts to derive a taxonomy of evaluation dimensions.* **Results and discussion:** *We identify four dimensions of evaluation of an elevator pitch: background and contextual knowledge showed in the presentation, project content and venture information, storytelling approach and style, and entrepreneurial flow or "algorithm" of the pitch. We detail the dimensions by defining 19 evaluation items and associated key evidence to support assessment. We undertake a preliminary application of the framework with three groups: Business investors, potential entrepreneurs, and entrepreneurship students.* **Conclusion and further research:** *We advance the discussion on venture storytelling and provide practitioners with a useful tool to support the evaluation of an entrepreneurial idea presented through an elevator pitch.*

Index Terms*—Business presentation, elevator pitch, entrepreneurial communication, evaluation, framework, technology entrepreneurship, venture creation.*

The process of exploiting emerging technologies to establish a new venture is commonly defined as technology entrepreneurship [1], [2]. It includes an articulated set of "desk" (idea stage), "premarket," and "market" activities, which have been substantially impacted by the emergence of digital technologies [3], [4], and by a new logic of entrepreneurial collaboration based on collective intelligence and crowdsourcing [5].

The complexity of the process, along with above-average risks and requirements of technological development, makes a particularly challenging endeavor for entrepreneurs. To raise early stage capital to transform the concept into a sustainable venture, and become "investment ready," the potential entrepreneur needs to interact with a number of potential stakeholders to whom the business idea or project is presented. The study of investors' decision parameters in the evaluation of a new venture [6]–[9] and the analysis of how entrepreneurs' social competence and business networks are positively related to financial success [10], [11] are not new. Fraiberg [12] has conceptualized entrepreneurs as workers able to deal with intricate interconnections among "genres, modes, languages, and spaces." Because failure to sell the idea and the opportunity may raise doubts in potential investors about the viability of the venture, entrepreneurs need to develop an "attention contract" [13] and advance their "impression management skills" so that the audience draws positive conclusions [14], [15].

In the idea-sharing process, the potential entrepreneur commonly delivers an elevator pitch—a concise, carefully planned, well-practiced, and business-oriented verbal message about the entrepreneurial idea. The name is derived from the supposition that an entrepreneur sharing an elevator ride with a potential investor has very little time to share the story and obtain a follow-up meeting. The term is used to indicate metaphorically an eye-catching presentation able to attract attention and interest, and the concept is

Manuscript received February 1, 2021; revised April 4, 2021; accepted April 5, 2021. Date of publication October 13, 2021; date of current version November 24, 2021. *(Corresponding author: Alessandro Margherita.)*
Alessandro Margherita is with the Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy (email: alessandro.margherita@unisalento.it).
David Verrill is with the Initiative on the Digital Economy, MIT Sloan School of Management, Cambridge, MA 02142-1347 USA (email: dverrill@mit.edu).

IEEE 10.1109/TPC.2021.3110620

Practitioner Takeaway

- Entrepreneurial storytelling and structured pitches are crucial to attract stakeholder interest and obtain potential commitment in technology ventures. Nevertheless, an integrative literature-based definition of the quality and evaluation pillars of an elevator pitch is still not available for practitioners.
- We present an integrative framework of four dimensions of evaluation—background, venture information, presenter style, and entrepreneurial "algorithm"—and detail the dimensions by defining 19 evaluation items and associated key evidence to support assessment. We undertake a preliminary application of the framework with early-stage investors, technology entrepreneurs, and entrepreneurship students.
- We advance the discussion on venture storytelling and provide practitioners with a useful tool to support the preparation and evaluation of an effective elevator pitch.

thus relevant in both business and nonbusiness situations. For example, in a *Nature* article, Doudna advocated for better training for future scientists, including guidance on how to craft a research elevator pitch [16].

A good elevator pitch is a combination of social skills, technical awareness, speed, and pragmatism required to influence the audience, and the use of a style appropriate to the nature and style of the project [17]. A study by van Werven et al. [18] has provided a comprehensive typology of arguments that an entrepreneur can use to obtain legitimacy, and a large stream of research [10], [14], [19]–[21] has evaluated the effect of entrepreneurs' presentation and communication skills on investors' decisions. Varga [22] analyzed aspects related to passion, audience knowledge, clarity of objectives, attire, and focus on benefits, and Diez-Prados [23] analyzed verbal and nonverbal engagement.

Although several guides on how to conduct a pitch are available in practice [24]–[29], scholars have typically focused on specific dimensions, such as defining different pitching strategies in different types of pitches [30]; the perceived entrepreneur's passion, preparation, and presentation design [31]; and the role of emotional and cognitive narratives [32]. However, a comprehensive framework of quality and evaluation dimensions of elevator pitches has not previously appeared. In particular, it would be relevant to integrate content and approach-related elements to propose a comprehensive idea of where a technology entrepreneurship project stands and what is its potential for success. Such an integrative effort would be relevant users, such as entrepreneurship students and potential entrepreneurs, technology entrepreneurs (TEs), and individuals engaged to assess the value of a new venture idea.

In our study, we focus in particular on entrepreneurs seeking funds and partners to launch their startup and the "preseed" pitches, longer than the typical short elevator pitch, which they present to reach the target audience. Such pitches are quite formalized in the practitioner experience in terms of parameters, such as length (about 10 minutes), structure, and visual aids used to convey the key messages. In most cases, these pitches are followed by a question-and-answer session aimed to deepen or clarify issues. We look at those pitches as presentation "archetypes" because they require (more than simple "idea disclosure" messages) a number of crucial requirements in terms of structure, content, and style or approach.

To achieve our research goals, we undertake a systematic review of extant literature to extract key concepts, ideas, and factors associated with the quality and effectiveness of an elevator pitch. We investigate business venturing, entrepreneurship, and business communication literature, looking also at systematization and practitioner-oriented contributions providing guides for building a successful venture and obtaining funds at the venture's early stages [33], [34]. We also looked at studies depicting the relation between idea and opportunity and the evolution of the venture from concept to exploitation (e.g., [35]), as well as studies highlighting the relevance of training programs to develop new entrepreneurs' ability to communicate and persuade the audience [36].

The next section of this article describes the method and process that we followed to

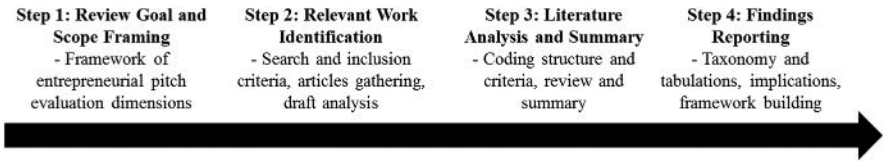


Fig. 1. Systematic review process.

systematically review the literature. Then, we illustrate the outcomes of that review and identify the specific dimensions and items of evaluation of an elevator pitch. Next, we reduce this conceptual exploration into a single analytical framework, which is also used for a preliminary application with three target user groups. Then, we discuss our findings and finally present conclusions, limitations, and paths for further research.

Research Process

We conducted a systematic review of specialized literature [37], [38] to gather definitions, conceptualizations, and analytical reviews of entrepreneurial presentations, with a specific focus on the elevator or business pitch. In particular, we conducted four activities, as shown in Fig. 1. In Step 1, we established the objective of the review, to derive an integrative framework, currently not present in the literature, of areas representing the perspectives of evaluation of an elevator pitch. The motivation and the boundaries of such framework development work were described in the previous section.

In Step 2, we used the Scopus database to search combinations of the keywords "elevator," "business," "entrepreneur∗" and "venture" with the keywords "pitch" and "speech" in article titles, abstracts, and keywords. The full search string used was *[TITLE-ABS-KEY ("elevator pitch") OR TITLE-ABS-KEY ("elevator speech") OR TITLE-ABS-KEY ("business pitch") OR TITLE-ABS-KEY ("business speech") OR TITLE-ABS-KEY ("entrepreneur∗ pitch") OR TITLE-ABS-KEY ("entrepreneur∗ speech") OR TITLE-ABS-KEY ("venture pitch") OR TITLE-ABS-KEY ("venture speech")]*. We limited the search to works published in journals to focus on academic contributions. We performed the query in mid-July 2020; the results are reported in Table I, divided by single keyword used. In fact, we also performed single searches to have an idea of the significance and diffusion of each concept. Due to duplications of works that can be found using different keywords, the total number of papers reported in Table I (88) is higher than the consolidated number of papers found using the full search string (78).

TABLE I
Number of Research Articles Found for Each Keyword

Keyword	Number of Articles
"elevator pitch"	34
"entrepreneur* pitch"	17
"business speech"	14
"business pitch"	11
"elevator speech"	10
"venture pitch"	2
Total (Consolidated Total)	**88 (78)**

The predominant practitioner focus on the topic (which is where our study aims to fill a gap by developing an academic investigation and integrative framework) has suggested that we be less restrictive in the initial article-gathering effort. We have extended the number and the scope of search keywords to retrieve a larger number of research articles and support a more inclusive conceptual extraction and systematization effort. Although concepts have been identified and used to assemble a larger number of studies to consider, we have carefully analyzed all the retrieved sources to check the appropriateness and real interest of each article for our purposes.

We reviewed the abstracts of the 78 papers and excluded works that are not relevant to our research (e.g., technical pitches in medical and biological fields) or that are generic business and entrepreneurship works not specifically focused on analyzing approaches, concepts, and processes related to making an entrepreneurial pitch. We thus selected 40 works, all included in the final reference list of this article. Table II lists the scientific journals where the selected articles were published. Most of those papers (23) were published in the *International Small Business Journal* (5 articles), *Journal of Business*

TABLE II
NUMBER OF PAPERS DIVIDED BY THE PUBLISHING JOURNAL

Journal	Number of Articles
International Small Business Journal	5
Journal of Business Communication	4
Entrepreneurship Theory and Practice	3
IEEE Transactions on Professional Communication	3
Academy of Management Journal	2
International Journal of Entrepreneurship and Innovation	2
Journal of Business Venturing	2
Journal of Research in Marketing and Entrepreneurship	2
Academy of Entrepreneurship Journal	1
Accounting Education	1
Business Communication Quarterly	1
Economic Research	1
Frontiers of Entrepreneurship Research	1
IEEE Transactions on Learning Technologies	1
International Journal of Entrepreneurial Behavior & Research	1
Journal of International Entrepreneurship	1
Journal of Management Development	1
Management Science	1
Organizational Behavior and Human Decision Processes	1
Proceedings of the National Academy of Sciences of the USA	1
Small Business Economics	1
Strategic Entrepreneurship Journal	1
Strategic Management Journal	1
Technical Communication Quarterly	1
Venture Capital	1
Total	**40**

Communication (4), *Entrepreneurship Theory and Practice* (3), IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION (3), *Academy of Management Journal* (2), *International Journal of Entrepreneurship and Innovation* (2), *Journal of Business Venturing* (2), and *Journal of Research in Marketing and Entrepreneurship* (2).

In Step 3, we analyzed the 40 selected papers to extract and annotate key concepts using a draft classification schema based on the main topic area. This extraction and aggregation work allowed us to gradually identify the following four main categories of concepts analyzed.

- Background knowledge and preparation required to hold a business pitch
- Presentation approach and personal style adopted by the presenter
- Content or information about the venture and the entrepreneur, which is shown in the pitch
- Logical reasoning or entrepreneurial "flow" of the pitch, and the entrepreneurial follow-up

Table III reports the outcome of the review and classification work, with the identification of the four areas of focus, the related concepts, and the associated references.

The next section describes Step 4 of the review process, in which we have used the retrieved items and key concepts to further detail the taxonomy of dimensions involved in the delivery and evaluation of the elevator pitch. For each of the four macroareas, we define a number of aspects and associated key points of analysis.

AREAS OF EVALUATION OF AN ELEVATOR PITCH

Background Knowledge and Preparation The delivery of an elevator pitch is based on preparation about the relevant technical background. The potential entrepreneur is expected to know what a company or business organization is and how it works, and what its distinguishing subsystems are (e.g., financial and technical resources, processes and operations, product and innovation outputs, and the financial and intangible value created). Besides, the realization of a technology-based business project also requires knowledge of project management methods and approaches. Previous research has addressed the multifaceted aspects and criteria involved with the evaluation of entrepreneurial speeches and storytelling. These include both dimensions mostly placed at the business (plan) level and aspects related to more "theatrical" aspects (e.g., 39, 40). Chen et al. [41] analyzed how a venture capitalist perceives entrepreneurial passion and how this perception may influence their investment decisions. That study showed that preparedness has a positive impact on funding decisions. Good preparation decreases stress, which can be heightened by an elevator pitch. In this regard, Burns [79] identified 7 Ps of speech anxiety reduction: *positive perception*, *positive thinking*, *positive self-talk*, *positive visualization*, *positive breathing*, *positive preparation*, and *positive behavior*.

A second area of "preparation" is related to the process required to build a technology venture. The roadmap to build a technology entrepreneurship venture includes a number of "idea," "premarket," and "market" stages [80], [81]. "Desk" activities focus on preliminary venture conceptualization and planning issues (i.e., opportunity definition, value proposition, and business planning). "Premarket"

TABLE III
FOCUS AREAS AND RELATED CONCEPTS

Focus Areas	Related Concepts (Alphabetical)	References
A. Background knowledge and preparation	• Adaptation to the target audience • Business preparation and background • Company processes and organization • Entrepreneurial preparation • Entrepreneurship education • Knowledge of the audience/stakeholders • Mastery of techno-entrepreneurship process • Organizational context of presentation • Preparation and anxiety reduction strategy • Preparation over passion • Stakeholder needs and expectations • Technology-based entrepreneurship training • Theatrical training and performance • Types of business presentation • Types of techno-preneurship projects	[39]-[46]
B. Presenter style and approach	• Body language and eye contact • Compellingness, credibility • Conciseness and clarity • Consistency • Discursive and linguistic characteristics • Displayed passion and enthusiasm • Dress code and business attire • Entrepreneur's coachability • Entrepreneurial pathos • Figurative language and gesturing • Informal approach to pitch • Multi-modal perspective • Oral presentation skills • Passion and investors decision • Persuasive non-verbal signs • Physical attractiveness • Presentational factors • Presenter enthusiasm • Role playing and empathy • Sensegiving • Signalled ethos • Skills, self-efficacy and relevance • Trustworthiness • Voice attractiveness	[19], [40], [41], [47]-[60]

activities include venture configuration and preparatory activities (i.e., capital raising, team organization, legal formation, and product development). Finally, "market" activities focus on venture implementation and process execution (i.e., management of operations, value appropriation, and venture growth). Although the practical experience is crucial for undertaking the activities above, research has shown that the profit of a new venture tends to be low when the entrepreneur has only startup, managerial, and high-growth experience without an educational background [82].

A third dimension of preparation or background-related aspects concerns issues, such as the

TABLE III
(*Continued*)

Focus Areas	Related Concepts (Alphabetical)	References
C. Venture information and pitch content	• Anecdotal openers and examples • Business model canvas • Business plan content • Cognitive legitimacy • Competitive space and positioning • Conceptual/functional traits of solution • Content reuse strategies • Customer touchpoints and channels • Discoursal template • Elements of the pitch • Entrepreneur's presentation • Entrepreneurial ties and network • Financial and revenue projections • Hybrid speeches • Key operations and technology • Neural engagement • Non-presentational factors • Organizational model and roles • Pricing strategy and mechanisms • Relevant need and call for action • Revenue sources • Sector or industry framing	[19], [61]-73]
D. Reasoning and logical flow of the pitch	• Assumptions and proof-of-concepts • Consistency of hypothesis • Feasibility and marketability • Entrepreneurial claims and micro-arguments • Entrepreneurship path • Pitcher commitment or "skin in the game" • Rational cognitive framework • Rationality over emotional aspects • Result, elaboration, purpose	[74]-[78], [46], [77]

purpose and types of business presentation, the required skills, and techniques for handling questions (e.g., TRACT: thank, repeat, answer, confirm, and thank). Delivering a business presentation or pitch requires playing a role, telling a story based on a canvas, and using voice and gestures to persuade the audience. Goffman's *The Presentation of Self in Everyday Life* [83] analyzes the theatrical performance involved in face-to-face interactions. When in contact with other people, an individual will attempt to guide the impression that others might make of them by changing setting, appearance, and behavior. This tendency applies also in entrepreneurial interaction, where the performer (potential entrepreneur) is on stage in front of the audience. Anderson and Warren extend Goffman's ideas by showing that "entrepreneurs have a distinctive presence in society that is shaped by cultural norms and expectations" [52, p. 589], whereas Anderson [40] studied the value of theatricality—particularly the application of dramatism and dramaturgy—as additional conceptual tools to analyze the entrepreneurial process.

Finally, the potential entrepreneur should exhibit contextual or situational awareness. The presenter should do a thorough analysis of the audience and the "environment" in which the pitch will be held. In particular, it is relevant to focus on values, needs, interests, constraints, backgrounds, and demographics of the audience. These elements will support the entrepreneur in the delicate effort of customizing the pitch and selecting the arguments

TABLE IV
ASPECTS RELATED TO BACKGROUND KNOWLEDGE AND PREPARATION

Aspect	Main Points
1. **Business management knowledge**	• Company as a system and its subsystems • Entrepreneurial education • Key company processes and organization • Basics of process and project management
2. **Technology entrepreneurship process**	• Peculiarities of technology entrepreneurship • Types of technology entrepreneurship projects • Desk, pre-market and market stage activities
3. **Foundations of business presentation and pitching**	• Meaning and purpose of a business presentation • Types of business presentation • Presentation profiles and skills, speaker's "function" • Definition and purpose of elevator pitch • Anxiety reduction strategies • Structure and techniques for holding a pitch
4. **Contextual awareness**	• Target audience composition, demographics • Stakeholder needs and expectations • Organizational context of presentation

used. Business planning literature highlights the importance of adapting the presentation—for example, by addressing the different focus of bankers (who stress financial aspects over the market and entrepreneur), versus venture capital and business angels (who highlight both market and financial dimensions). Business angels place more emphasis than venture capitalists on the entrepreneur and "investor fit" considerations [39]. Fisher et al. [44] showed that the criteria for a new technology venture may vary according to the audience. Table IV reports a snapshot of aspects related to the background and preparation dimensions, which were extracted from the literature and integrated with the introductory information provided earlier in this paragraph.

Presenter Style and Approach TEs pitch their business idea or venture with the goal of creating audience interest by offering credible and applicable evidence. Oral communication skills are thus crucial for a nascent entrepreneur. More generally, oral presentation and communication skills are also considered by managers among the most important competencies for graduates entering the workforce [49]. A pitch is more than a presentation transmitting information. It is more like a conversation that helps the listener to imagine how the world would be if the new idea being presented were realized. Commitment is expected to produce action by the audience. The elevator pitch can be stressful because it creates pressure on the presenter. The time is brief, and the presence of "judges" in a formal setting may hinder the real potential of the entrepreneur to tell the full story in an effective way. In this regard, the traditional elevator pitch could be replaced by a model of reception elevator pitch [57].

O'Leary [25] described the 9Cs of a perfect pitch. First, the pitch should be *concise*—succinct and clear, with as few words as possible. Second, the pitch should be *clear*—not filled with acronyms and technical jargon. Third, the pitch should be *compelling*—stating problems that can be solved and describing how they could be addressed. Fourth, the pitch has to be *credible*— spelling out what makes the entrepreneur qualified to do what is being envisioned. Fifth, the pitch is *conceptual*—initially staying at a high level. Sixth, the pitch is *concrete*—tangible and easily grasped. Seventh, the pitch has to be *consistent*—showing coherence and making a durable fit with the target audience. Eighth, the pitch is *customized*—adapted in style and content to the specific audience. Ninth, the pitch is *conversational*—capturing the target and stimulating discussion and understanding.

A second aspect of evaluation concerns the presenter. Presentation factors [19] relate to the entrepreneur's style of delivery. Entrepreneurs adopt different tactics to engage the audience.

TABLE V
ASPECTS RELATED TO PRESENTER STYLE AND APPROACH

Aspect	Main Points
1. **Stylistic points**	• Conciseness and information choice • Overall clarity and consistency • Compellingness, credibility, ability to involve • Conceptualization and concreteness • Customization and conversational style
2. **Presentation style**	• Verbal and narrative style • Gestures, voice, body language, eye contact • Dress code and business attire • General presenter's style and approach
3. **Emotional levers**	• Entrepreneurial pathos • Role playing and empathy with audience • Displayed passion, enthusiasm and coachability

Common tactics include demonstrations (i.e., live demos and photos of a new product or technology) and real-life stories. Clarke et al. [59] showed how figurative language and gestures may also influence investment decisions, since they help investors to enhance the perception of one venture's potential. Sensegiving may also play a direct role in achieving support for a venture [53]. In the early stages of a venture, metaphors are used to emphasize agency and control, and the predictability and transparency of a novel venture. It is also important to note the relevance of addressing cultural sensibilities (e.g., eye-to-eye contact or gestures).

Finally, enthusiasm and emotional layers are crucial aspects to consider. In the 1990s, Beck and Wegner [47] studied the role of enthusiasm in technical proposals and found differences between social science/humanities proposals and science/engineering ones. The concept of "signaled ethos" [48] has been discussed as a useful mechanism for understanding the "images" that authors create of themselves. There are five categories of signaled ethos in business speeches: "deference," "self-criticism," "similitude," "expertise," and the "inclination to succeed." Cardon et al. [50] showed that the display of entrepreneurial passion is a factor in business angels' investment decisions, which are shaped by the enthusiasm, preparation, and commitment that entrepreneurs exhibit. Table V shows the dimensions of pitch evaluation that relate to the presenter, which were extracted from the literature and integrated with the introductory information provided in this paragraph.

Venture Information and Pitch Content

Besides looking at how things are communicated, it is crucial to evaluate what is being said. Pollack et al. [66] have highlighted the positive correlations among entrepreneurs' preparedness, cognitive legitimacy, and the amount of funding received. Cognitive legitimacy is the knowledge about a new activity and what is needed to succeed [84]. Substance-oriented or nonpresentation criteria include company, market, product, and finance issues [19].

There are many business and entrepreneurship textbooks and practitioner guides [24], [85] that define the key elements to be included in a pitch. The pitch should start with a strong opening (or "hook") that grabs the audience's attention. An introduction increases the audience's willingness to listen and understand a speech. In particular, anecdotal openers (or personas) increase the presentation's comprehensibility and the speaker's credibility [62].

Spinuzzi et al. [67] studied strategy reuse in entrepreneurs' pitches. Entrepreneurs reuse professional communication content, and this reuse can be characterized as accepting (repeating verbatim), continuing (extending arguments), and resisting (rebutting arguments). Kirsch et al. [63] showed that venture capital funding decisions are weakly associated with the preparation and submission of planning documents and information.

The pitch should state an unmet customer need (problematic scenario) and explain how the

TABLE VI
ASPECTS RELATED TO VENTURE INFORMATION AND PITCH CONTENT

Aspect	Main Points
1. Contextual Introduction	• Anecdotal openers and examples • Techno-human-social endeavour • Sector or industry framing • Big trend or challenge faced
2. Customer Pain	• Relevant need, market gap or call for action • Problem characteristics and intensity • Explicit or latent call for action
3. Solution	• Conceptual/functional traits of solution • Competitive space and positioning • Response to extant gap and product
4. Team and Roles	• Current board and team members • Motivation, education, experience • Organizational model and roles
5. Business Process	• Key operations and technology • Value chain and value system • Customer touchpoints and channels
6. Revenue and Financials	• Pricing strategy and mechanisms • Revenue sources • Financial and revenue projections
7. Ask and Follow-up	• Timeline and milestones • Pre and post-money evaluation • Financial and/or advisory demand

proposed solution can satisfy that need. The presentation should leave the audience wanting to know more, and it should be concluded with an explicit request ("ask") for financial, technical, network, or expertise assistance (call for action).

First, the contextual opening should be framed in terms of a techno-human-social endeavor, sector/industry framing, and trend or challenge. Second, customer pain includes the relevant need or market gap, problem characteristics/intensity, and call for action. Third, the solution includes the conceptual/functional solution, the competitive space and mapping, the response to the gap, and the product. Fourth, the team and roles include the current board, advisors, employees (or team), motivation, education, experience, organizational model, and roles. In this aspect, it is relevant to highlight the role of "entrepreneurial ties" to increase the innovativeness of the startup, and the positive impact in terms of expected level of competition, novelty to customers, and novelty of technology [65]. Fifth, the business process describes key operations and technology, the venture value system, and customer channels. Sixth, revenue and financials include pricing strategy and mechanisms, revenue sources, and financial projections. Seventh, the ask and follow-up include the timeline and milestones, pre- and postmoney valuation, and financial and advisory needs. Table VI provides a list of evaluation topics and specific items related to the venture, which were extracted from the literature and integrated with the introductory information provided in this paragraph.

Reasoning and Logical Flow of the Pitch

Although the entrepreneur's approach, style, and emotional aspects are crucial to convince the audience in an entrepreneurship presentation, there is an important link between the propensity to fund a venture and the presence of rational aspects in the pitch. In fact, although emotional appeals are useful to reinforce arguments, they are not sufficiently persuasive on their own [76]. The

lack of rational elements may thus guarantee the failure of the entrepreneur's attempt to be persuasive.

Beyond the overall structure and content of the entrepreneurial narrative, it is thus crucial that the pitch present solid arguments that can help achieve resonance and plausibility. In this regard, van Werven et al. [78] identify the claims, arguments, and rhetorical strategies that entrepreneurs adopt when holding a pitch. The claims are related to the following.

- Customer needs and call for action
- Product/solution and its ability to provide benefits
- Venture and enabling environment
- Intermediate or expected performance of the venture
- Team composition and capabilities
- Funding and other requests

The elevator pitch should represent and demonstrate the existence of a consistent logical flow, a robust stream of ideas, which we call the "entrepreneurial algorithm." An algorithm is a method or process for solving a problem or accomplishing an objective through a number of steps, and the creation of a technology venture is a process for addressing a customer problem and accomplishing the entrepreneur's business objectives. The pitch is thus aimed at describing the process and demonstrating that it can really work, starting from an "input"—a problem, interest, vision, passion, or need of the TE—and ending with an "output" —a venture that exists and is sustainable on the market. A precondition or initial assumption is that the problem, although complex and challenging, can actually be solved.

Then, there are four "assessment points" that build up the "entrepreneurial algorithm," which are as follows.

- Existence of a concrete solution in the hands (or in the mind) of the TE
- Realizability of the solution by the TE and his/her team
- Identification of a significant market potential, with clear customers segments, for the solution
- Current feasibility of the venture in terms of process and financial maturity

Fig. 2 shows the entrepreneurial algorithm, with the four assessment points and the negative "ways out"—the solution gap, realization gap, market gap, and preparation gap. The chart, which addresses one of the four dimensions of the proposed framework—the logical flow or reasoning "algorithm" underlying the pitch—is aimed at showing the key evaluation (and self-assessment) steps that can be identified as a go/no go pitch.

Elevator Pitch Assessment Model (EPAM)

Assessment Scorecard In the previous sections, we have used the outcomes of the literature review to identify four areas of evaluation of an elevator pitch. For each area, we have developed a schema to describe a number of specific items. Building on those findings, this section presents an integrative assessment scorecard or EPAM. Table VII shows the scorecard with the four dimensions of evaluation and the detail of each dimension with the identification of 19 specific evaluation items. For each item, Table VII also reports evidence or "signals" that can support the assessment or the stakeholder's reflection on that aspect.

The EPAM is specifically focused on preseed pitches held by entrepreneurs in front of potential investors and partners of their startup. This form of entrepreneurial storytelling is formalized in the practitioner experience in terms of a number of aspects, such as length, structure, and visual aids, used to convey the key messages. Although the application of the model to other business or nonbusiness situations should be assessed before it is applied (different audiences and situations require different content and approaches), the framework can be potentially applied to pitches held in front of a larger range of relevant stakeholders besides investors and partners, such as customers, distributors, and potential team members.

Note that the framework is not an assessment or evaluation tool per se, and more fine-grained evaluation criteria could be identified. But the dimensions of analysis reported in Table VII provide a checklist of parameters to be operationalized using more specific measurement scales and parameter.

Preliminary Applications with Target Users From a practitioner perspective, the framework has the following three potential types of users.

1. The assessment model can be used by **TEs** engaged to deliver a pitch in different phases of the venture creation journey. The framework offers these users a comprehensive checklist of issues to address to deliver an effective pitch in

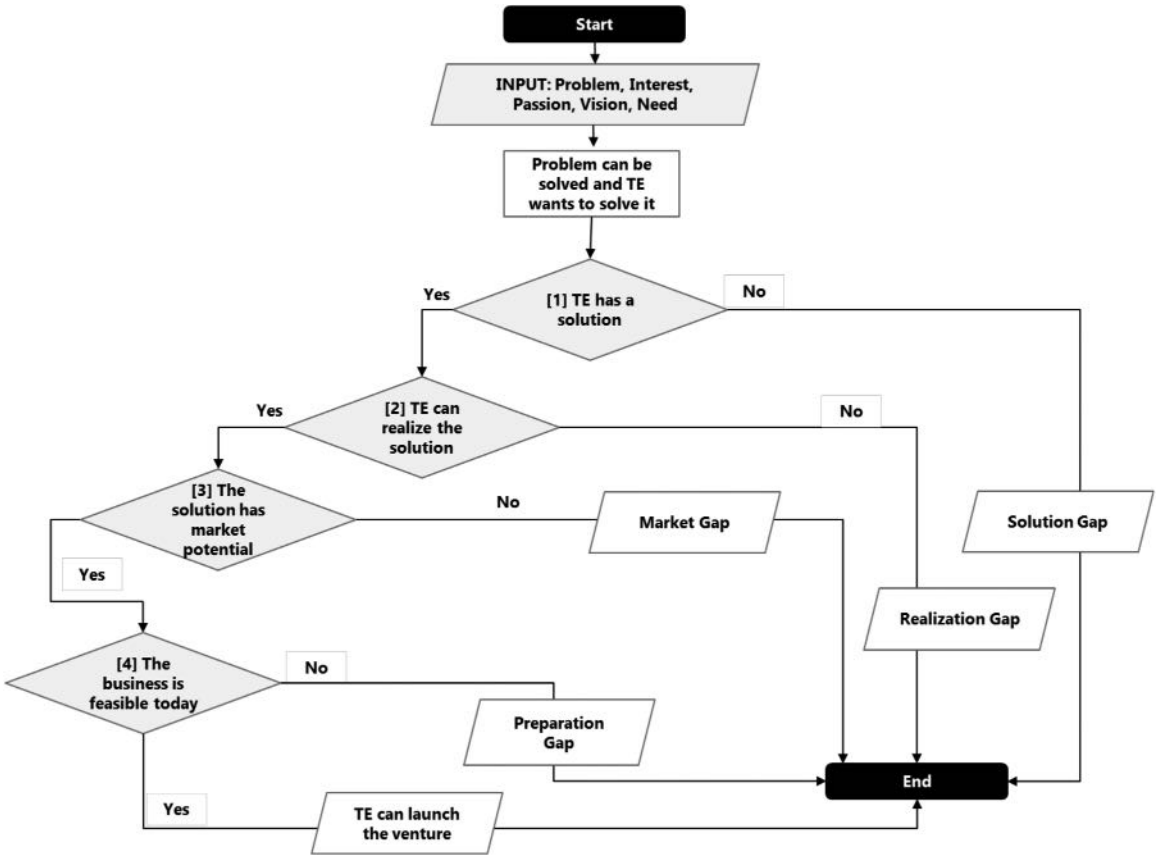


Fig. 2. Reasoning and logical flow of the pitch.

front of a number of entrepreneurial stakeholders.

2. Another potential user of the model is **entrepreneurship students** involved in business communication and elevator pitch training sessions. Our model can be used as an education tool to support the development of purposeful competencies and skills. At the same time, the model can be useful for educators engaged to evaluate students' presentations and elevator pitches, which are delivered in coherence with the provided model.
3. Finally, the model can be useful for **investors** to support their venture evaluation and investment decision process. Although the investment decision is largely dependent on other aspects beside the pitch, the framework provides a literature-based model of evaluation aspects that can be useful to define a comprehensive assessment.

We conducted a preliminary application of the framework with each of these user types.

First, we submitted the model for an expert review to **two investments entities**: a Boston-based angel group and an Italian business angel. We sent the elevator pitch assessment scorecard (see Table VI) to the managing directors of these investment companies and asked them to review the model to determine whether all the aspects have been properly identified and described. The managers provided general feedback on the framework, based on test applications of the framework in entrepreneurial meetings and evaluation sessions, along with amendments and suggestions for refinement.

A second application of the model was done in an engineering course for 25 undergraduate **students** focused on technological innovation and technology-based entrepreneurship. The course, held at the Department of Engineering for Innovation of the University of Salento, included about 80 hours of mixed theory development and practical training. A part of the learning process was dedicated to training the students to prepare and deliver an elevator pitch for a technology venture project. Fostering an entrepreneurial mindset and intentions, and networking skills in students is indeed a critical function of modern entrepreneurship programs [42], [86]–[88]. The

TABLE VII
ELEVATOR PITCH ASSESSMENT MODEL

Dimension	Evaluation Items	Key Evidence
1. **Background and Contextual Knowledge Emerging from the Pitch** (background-related dimension)	**1.1** Knowledge of nature/characteristics of technology entrepreneurship and phases/activities of the technology venturing process	• Idea/project is properly positioned in a venturing roadmap • Presenter shows awareness of risks of techno-entrepreneurship
	1.2 Knowledge of a business organization, its subsystems and critical components and dynamics or process mechanisms	• Clear business jargon and semantic embedded in narration • Presenter frames a relevant organizational scenario
	1.3 Knowledge of audience to meet, including demographics and professional profile and expectations, as well as of the physical/organizational context of presentation	• Information and messages are customized to audience • Presenter shows mature situational awareness
	1.4 Knowledge of methods, practices and approaches to delivering an effective business presentation, with a specific focus on business and entrepreneurial pitches	• Self-confidence and professional etiquette is showed • Presenter masters business presentation standards
2. **Project Content and Venture Information Shared During the Pitch** (venture-related dimension)	**2.1** Information related to the problematic scenario and context of reference of the business idea or venture project	• Evidence of data/facts from authoritative sources is brought • Presenter knows trends and uses role-playing
	2.2 Information on specific customer pains and unmet needs showed by a reasonable population of people which can be potentially served	• Order of magnitude of market segments is provided • Presenter is aware of likely evolution of the customer base
	2.3 Information on the conceptual, functional and technological key points of the solution, along with the value added respect to extant offers	• Core product/service layout is showed • Presenter knows/mentions direct and indirect competitors
	2.4 Information on the presenter and the team of people who will realize or support realization of the solution	• Organization model for the venture is sketched • Presenters explains roles, track record and complementarities
	2.5 Information on the business process to realize the product/service, with positioning/specialization of the venture/team	• Value chain/system model of the company is described • Presenter shares make or buy and outsourcing strategy
	2.6 Information on key figures related to sales, revenues/costs and overall profitability estimates of the project on a three years' timeframe	• Break-even point and payback estimates are provided • Presenters anticipates return on equity/investment
	2.7 Information on current needs and resource gaps of the venture team, which drives a call for participation of the audience of stakeholders	• Formulation of a clear, direct and justified request • Presenter shares rules of engagement and scenarios

"camp model" described in entrepreneurship education is focused on the generation or transformation of new ideas in concepts and basic plans [89]. In our application, we required students to prepare their pitches using the model as a design and evaluation scorecard, which was subsequently used by the instructors to define improvements and annotations on the presented pitch, and to formulate an evaluation.

Finally, we provided the model to three **TEs** engaged in product development and partner scouting activities. The entrepreneurs, working on a digital publishing project, a multiservice wine selling platform, and a digital-enhanced accommodation service, have interacted with one of the authors of this article for a number of mentoring sessions. The entrepreneurs were thus provided with the framework to support the preparation of business presentations and elevator pitches held in June and July 2020. Table VIII shows the three applications of the model, with details of the participants, locations, application

TABLE VII
(*Continued*)

Dimension	Evaluation Items	Key Evidence
3. Storytelling Approach and Style Qualities of the Pitch (presenter-related dimension)	**3.1** Quality of the entrepreneurial narration as related to clarity, synthesis, concreteness, compellingness, ability to involve	• Narration is overall enjoyable and balanced • Presenter captures audience and nurtures interest
	3.2 Quality of the verbal approach adopted as related to voice speed, volume, emphasis, variety, and intensity	• Overall verbal approach is well prepared and professional • Presenter modulates voice to highlight ideas and concepts
	3.3 Quality of the non-verbal communication are related to body language, gestures and eye contact	• Non-verbal approach enhances storytelling • Presenter has authoritative and professional presence
	3.4 Quality of emotional intensity during narration, emerging entrepreneurial passion, pathos, and courage	• Narration attracts, involved and strikes audience • Presenter shows personal commitment in the project
4. Logical Flow and Entrepreneurial Algorithm in the Pitch (flow-related dimension)	**4.1** Demonstration that there is a consistent technology-based solution to a clear, timely and relevant problem	• A timely technology trend and social issue are addressed • Presenter explain the links between problem and solution
	4.2 Demonstration that the solution can be realized (by a socio-technical point of view) in practice and can have an impact on industry	• Engineering design of the solution is showed • Presenter presents or shares a prototypal version
	4.3 Demonstration that the solution can generate products and services with an actual market which is sustainable and scalable in the long term	• Rough pricing and revenue generation are discussed • Presenter presents market/sales projections and estimates
	4.4 Demonstration that the realization of solution and the underlying business process are in line with the entrepreneur/team profile and ability to attract support	• Track record of the presenter and team's potential is discussed • Presenter explains how gaps can be filled effectively

processes (information collection and interaction), and feedback.

These applications provided insights on the utility of the model as an integrative checklist to support the preparation of an elevator pitch or business meeting (entrepreneur and student perspective) or to evaluate the potential of an entrepreneurial project (investor and trainer perspective). The model can thus be a tool for those engaged in venture assessment and entrepreneurial evaluations. Two potential critical factors are that the number of items in the scorecard makes it complex to use in real time and that investment decisions are of course based on aspects beyond the business presentation (due diligence, personal knowledge, and other extra-presentation factors). Students found the scorecard useful to know in advance the criteria used to evaluate their work (process transparency) and the key points to focus on to improve the pitch. One point raised by the students is the need to accompany the scorecard with a more detailed guide or handbook. Finally, the framework provides useful insights on investors' expectations and required evidence, and it can be a useful template for structuring a business presentation. However, customized versions of the model for specific audiences would be useful.

Although the application in different contexts may be constrained in terms of generalizability, the model offers an integrative definition of key general evaluation dimensions and items, which can be useful in different stakeholder contexts, with the real customization need being in terms of the choice and depth of information to convey to the specific audience.

Discussion

Launching a technology venture is based on an articulated array of activities in which the potential entrepreneur interacts with several stakeholders

TABLE VIII
APPLICATIONS OF THE EPAM AND KEY FEEDBACK OBTAINED

Type of Application	Details	Key Feedback
1. **Investor**	• Participants: 2 investment companies • Location: USA/Italy • Interaction method: Video meetings and e-mails • Application process: provision of the scorecard (EPAM) to investors, use of EPAM as a systematization checklist to evaluate a new venture, gathering of participant feedback	• The scorecard provides a comprehensive analysis of what constitutes a professional elevator pitch • The tool can be used as practical checklist to support assessment or entrepreneurial evaluations • The framework could be of potential complexity for application in real-time (during the pitch itself) • The investment decision is not only based on the pitch but also on extra-presentation aspects
2. **Student**	• Participants: 25 undergraduate students • Location: Italy • Interaction method: Face-to-face • Application process: provision of the scorecard (EPAM) to students, use of EPAM as a canvas to prepare and deliver the pitch, gathering of participant feedback	• The early knowledge of the framework ensures transparency of evaluation criteria • The scorecard is a useful checklist for the design, preparation and evaluation of the pitch • The model should be supported or accompanied with practical guides which explain the "how-to"
3. **Entrepreneur**	• Participants: 3 technology entrepreneurs • Location: Italy • Interaction method: Face-to-face and video- meetings • Application process: provision of the scorecard (EPAM) to entrepreneurs, use of EPAM to support preparation of meetings with potential partners, gathering of participant feedback	• The framework provides useful insights on investors' expectations and required evidence or information • The scorecard is a useful template for structuring the business presentation • More customized versions of the model would be useful for different stakeholders

who may support idea validation, construction, and execution. To convince potential investors, the entrepreneur normally uses a business or elevator pitch to share the value proposition and the deriving opportunities for engagement, establishing with the audience a sort of platform for communicative interaction [90]. The quality of the elevator pitch and its potential to persuade stakeholders to get involved in the entrepreneurial projects depend on a number of factors related to the presenter, the venture idea, and the narrative provided.

In this article, we have addressed the lack of a theory-based and integrative definition of these factors, and we have presented a systematic analysis of key evaluation dimensions affecting the quality and effectiveness of an elevator pitch, and the consequent evaluation of the potential new venture. We aim to propose a pragmatically oriented design [91] for evaluating entrepreneurial presentations and the potential of new venture ideas. The framework includes four areas related to background knowledge and preparation, pitching approach and style, content of the narration, and the logical flow of arguments made in the presentation. We have detailed these four areas by defining 19 items and related evidence that can support the assessment of an entrepreneurial pitch. We also validated our framework with two early stage investment experts, three TEs, and a group of 25 students.

Our research contributes to existing theory on entrepreneurship and business venturing by providing a framework that integrates a large but fragmented academic and practitioner knowledgebase on what constitutes a good entrepreneurial pitch. First, we complement previous attempts to identify success factors and drivers of positive investors' decisions, which have been focused on specific emotional aspects [32], [76] or the use of speech and gestures to influence the audience [53]. In our article, the emotional

dimension is incorporated in the presenter-related aspects, which are complemented by three other types of quality drivers for an entrepreneurial pitch.

We found significant similarities and differences between our work and that of van Werven et al. [78]. In particular, although the focus and the nature of our findings are quite dissimilar, we share an attempt to provide an integrated framework to support further theory development and a relevant guidance for practitioners as well. The article by van Werven et al. [78] describes high-level claims that entrepreneurs make in their presentations, discounts claims by identifying underpinning or supporting arguments, and associates a number of rhetorical strategies. Claims and arguments are related to customer needs and the call for solution, the product and its benefits, the environment of the venture, the current and expected performance, the team's composition and skills, and the funding needs or other "asks" directed to the audience. That study is thus the systematization of a good narrative along with claims and rhetorical strategies. Our work expands that contribution by identifying four areas of discussion in which claims are embedded. In particular, the content of claims and the rhetorical strategies are elements that are found in the content-related dimensions and in the entrepreneurial algorithm of our framework. Moreover, we also provide details and evidence for each item associated with the four areas.

Conclusion

The creation of a new technology-based venture brings the potential entrepreneur to interact with a number of different actors and persuade them of the quality of the business project. In such an endeavor, entrepreneurial storytelling and business pitches are crucial to attract stakeholder interest and potential commitment. In this article, we systematically review the literature and identify 40 research articles from which we derive concepts associated with four areas that impact the quality and effectiveness of an elevator pitch. We further detail these areas by classifying 19 specific items and evidence types to support pitch delivery and evaluation, and undertake a preliminary application with three potential users of the framework.

This article has several limitations, which can also represent paths for future research. First, the model was validated only by submitting the framework to a single-validation process, although it is articulated in three contexts. A more extended scholarly evaluation—such as including technology entrepreneurship and business communication experts, and the application of real business venturing analysis—would support the definition of a more robust model. Second, although the scorecard is useful for evaluating an elevator pitch, its operationalization to support practical preparation and evaluation would require development of measurable items for each evaluation aspect of the scorecard.

References

[1] D. Jones-Evans, "A typology of technology-based entrepreneurs: A model based on previous occupational background," *Int. J. Entrepreneurial Behav. Res.*, vol. 1, no. 1, pp. 26–47, 1995.
[2] T. H. Liu, S.-C. Hung, S.-Y. Wu, and Y.-Y. Chu, "Technology entrepreneurial styles: A comparison of UMC and TSMC," *Int. J. Technol. Manage.*, vol. 29, no. 1/2, pp. 679–683, 2005.
[3] G. Elia and A. Margherita, "Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process," *Technol. Forecasting Social Change*, vol. 150, 2020, Art. no. 119791.
[4] S. Kraus, C. Palmer, N. Kailer, F. L. Kallinger, and J. Spitzer, "Digital entrepreneurship: A research agenda on new business models for the twenty-first century," *Int. J. Entrepreneurial Behav. Res.*, vol. 25, no. 2, pp. 353–375, 2019.
[5] V.V. De Luca, A. Margherita, and G. Passiante, "Crowdfunding: A systemic framework of benefits," *Int. J. Entrepreneurial Behav. Res.*, vol. 25, no. 6, pp. 1321–1339, 2019.
[6] I. C. MacMillan, R. Siegel, and P. N. Subba Narisimha, "Criteria used by venture capitalists to evaluate new venture proposals," *J. Bus. Venturing*, vol. 1, no. 1, pp. 119–128, 1985.
[7] R. D. Hisrich and A. D. Jankowitz, "Intuition in venture capital decisions: An exploratory study using a new technique," *J. Bus. Venturing,* vol. 5, no. 1, pp. 49–62, 1990.
[8] V. H. Fried and R. D. Hisrich, "Toward a model of venture capitalist decision-making," *Financial Manage.*, vol. 23, no. 3, pp. 28–37, 1994.
[9] J. Hall and C. W. Hofer, "Venture capitalists' decision criteria in new venture evaluation," *J. Bus. Venturing*, vol. 8, no. 1, pp. 25–42, 1993.
[10] R. A. Baron and G. D. Markman, "Beyond social capital: The role of entrepreneurs' social competence in their financial success," *J. Bus. Venturing,* vol. 18, no. 1, pp. 41–60, 2003.

[11] H.-C. Huang, M.-C. Lai, and K.-W. Lo, "Do founders' own resources matter? The influence of business networks on start-up innovation and performance," *Technovation*, vol. 32, no. 5, pp. 316–327, 2012.
[12] S. Fraiberg, "Start-up nation: Studying transnational entrepreneurial practices in Israel's start-up ecosystem," *J. Bus. Tech. Commun.*, vol. 31, no. 3, pp. 350–388, 2017.
[13] D. Valliere, "Entrepreneurial sensegiving and the attention contract," *Int. Entrepreneurship Manage. J.*, vol. 11, no. q, pp. 77–94, 2015.
[14] C. M. Mason and R. T. Harrison, "Auditioning for money: What do investors look for at the initial screening stage?" *J. Private Equity*, vol. 6, no. 2, pp. 29–42, 2003.
[15] A. Parhankangas and M. Ehrlich, "How entrepreneurs seduce business angels: An impression management approach," *J. Bus. Venturing*, vol. 29, no. 4, pp. 543–564, 2014.
[16] J. Doudna, "Genome-editing revolution: My whirlwind year with CRISPR," *Nature*, vol. 528, no. 7583, pp. 469–471, 2015.
[17] S. Manning and T. A. Bejarano, "Convincing the crowd: Entrepreneurial storytelling in crowdfunding campaigns," *Strategic Organiz.*, vol. 15, no. 2, pp. 194–219, 2017.
[18] R. van Werven, O. Bouwmeester, and J. P. Cornelissen, "The power of arguments: How entrepreneurs convince stakeholders of the legitimate distinctiveness of their ventures," *J. Bus. Venturing,* vol. 30, no. 4, pp. 616–631, 2015.
[19] C. Clark, "The impact of entrepreneurs' oral "pitch" presentation skills on business angels initial screening investment decisions," *Venture Capital*, vol. 10, no. 3, pp. 257–279, 2008.
[20] M. N. Hoehn-Weiss, C. G. Brush, and R. A. Baron, "Putting your best foot forward? Assessments of entrepreneurial social competence from two perspectives," *J. Private Equity*, vol. 7, no. 4, pp. 17–26, 2004.
[21] P. Davidsson and B. Honig, "The role of social and human capital among nascent entrepreneurs," *J. Bus. Venturing*, vol. 18, no. 3, pp. 301–331, 2003.
[22] S. Varga, *Brilliant Pitch: What to Know, Do and Say to Make the Perfect Pitch*. New York, NY, USA: Pearson Education, 2009
[23] M. Díez-Prados, "Verbal and nonverbal engagement devices in business persuasive discourse: The elevator pitch," in *Engagement in Professional Genres*. C. Sancho Guinda, Ed. Amsterdam, The Netherlands: John Benjamins, 2019.
[24] A. Pincus, "The perfect (elevator) pitch," *Bloomberg Bus. Week*, Jun. 18, 2007.
[25] C. O'Leary, *Elevator Pitch Essentials: How to Get Your Point Across in Two Minutes or Less*. Saint Louis, MO, USA: The Limb Press LLC, 2008.
[26] O. Klaff, *Pitch Anything: An Innovative Method for Presenting, Persuading, and Winning the Deal*. New York, NY, USA: McGraw-Hill, 2011.
[27] R. Porter, "Crafting a sales pitch for your grant proposal," *Res. Manage. Rev.*, vol. 18, no. 2, pp. 79–84, 2011.
[28] P. Coughter, *The Art of the Pitch: Persuasion and Presentation Skills That Win Business*. London, UK: Palgrave Macmillan, 2012.
[29] S. W. Martin, "Win the business with this elevator pitch," *Harvard Bus. Rev. Blog*, Aug. 22, 2012.
[30] B. Teague, M. D. Gorton, and X. X. Liu, "Different pitches for different stages of entrepreneurial development: The practice of pitching to business angels," *Entrepreneurship Regional Develop.*, vol. 32, no. 3/4, pp. 334–352, 2020.
[31] C. S. Galbraith, B. C. McKinney, A. F. DeNoble, and S. B. Ehrlich, "The impact of presentation form, entrepreneurial passion, and perceived preparedness on obtaining grant funding," *J. Bus. Tech. Commun.*, vol. 28, no. 2, pp. 222–248, 2014.
[32] A. Wuillaume, A. Jacquemin, and F. Janssen, "The right word for the right crowd: An attempt to recognize the influence of emotions," *Int. J. Entrepreneurial Behav. Res.*, vol. 25, no. 2, pp. 243–258, 2019.
[33] S. Blank and B. Dorf, *The Startup Owner's Manual: The Step-by-Step Guide for Building a Great Company*. Pescadero, CA, USA: K&S Ranch, 2012.
[34] G. Daugherty, *Startup Success: Funding the Early Stages of Your Venture*. Austin, TX, USA: River Grove Books, 2019.
[35] P. Vogel, "From venture idea to venture opportunity," *Entrepreneurship Theory Pract.*, vol. 41, no. 6, pp. 943–971, 2017.
[36] C. Spinuzzi, D. Altounian, and G. P. Pogue, "Go or no go: Learning to persuade in an early-stage student entrepreneurship program," *IEEE Trans. Prof. Commun.*, vol. 63, no. 2, pp. 100–117, Jun. 2020.
[37] J. F. Wolfswinkel, E. Furtmueller, and C. P. Wilderom, "Using grounded theory as a method for rigorously reviewing literature," *Eur. J. Inf. Syst.*, vol. 22, no. 1, pp. 45–55, 2013.
[38] D. Denyer and D. Tranfield, "Producing a systematic review," in *The SAGE Handbook of Organizational Research Methods*. D. A. Buchanan and A. Bryman, Eds. London, UK, Sage: 2009, pp. 671–689.
[39] C. M. Mason and M. Stark, "What do investors look for in a business plan? A comparison of the investment criteria of bankers, venture capitalists and business angels," *Int. Small Bus. J.*, vol. 22, no. 3, pp. 227–248, 2004.
[40] A. R. Anderson, "Enacted metaphor: The theatricality of the entrepreneurial process," *Int. Small Bus. J.*, vol. 23, no. 6, pp. 587–603, 2005.
[41] X.-P. Chen, X. Yao, and S. Kotha, "Entrepreneur passion and preparedness in business plan presentations: A persuasion analysis of venture capitalists' funding decisions," *Acad. Manage. J.*, vol. 52, no. 1, pp. 199–214, 2009.
[42] T. J. Bae, S. Qian, C. Miao, and J. O. Fiet, "The relationship between entrepreneurship education and entrepreneurial intentions: A meta–analytic review," *Entrepreneurship Theory Pract.*, vol. 38, no. 2, pp. 217–254, 2014.

[43] J. Schneider, D. Börner, P. Van Rosmalen, and M. Specht, "Can you help me with my pitch? Studying a tool for real-time automated feedback," *IEEE Trans. Learn. Technol.*, vol. 9, no. 4, pp. 318–327, Oct.–Dec. 2016.
[44] G. Fisher, D. F. Kuratko, J. M. Bloodgood, and J. S. Hornsby, "Legitimate to whom? The challenge of audience diversity and new venture legitimacy," *J. Bus. Venturing*, vol. 32, no. 1, pp. 52–71, 2017.
[45] D. Clingingsmith and S. Shane, "Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States," *Manage. Sci.*, vol. 64, no. 11, pp. 5164–5179, 2017.
[46] A. M. Ducasse, "Evidence-based persuasion: A cross-cultural analysis of entrepreneurial pitch in English and Spanish," *J. Int. Entrepreneurship*, vol. 18, no. 4, pp. 492–510, 2020.
[47] C. E. Beck and K. A. Wegner, "Enthusiasm in technical proposals: Verifying a method of lexical analysis," *IEEE Trans. Prof. Commun.*, vol. 33, no. 3, pp. 118–123, Sep. 1990.
[48] L. Beason, "Strategies for establishing an effective persona: An analysis of appeals to ethos in business speeches," *J. Bus. Commun.*, vol. 28, no. 4, pp. 326–346, 1991.
[49] T. G. Weldy and M. L. Icenogle, "A managerial perspective: Oral communication competency is most important for business students in the workplace," *J. Bus. Commun.*, vol. 34, no. 1, pp. 67–80, 1997.
[50] M. S. Cardon, R. Sudek, and C. Mitteness, "The impact of perceived entrepreneurial passion on angel investing," *Frontiers Eentrepreneurship Res.*, vol. 29, no. 2, pp. 1–15, 2009.
[51] D. Cyphert, "The rhetorical analysis of business speech: Unresolved questions," *J. Bus. Commun.*, vol. 47, no. 3, pp. 346–368, 2010.
[52] A. R. Anderson and L. Warren, "The entrepreneur as hero and jester: Enacting the entrepreneurial discourse," *Int. Small Bus. J.*, vol. 29, no. 6, pp. 589–609, 2011.
[53] J. P. Cornelissen, J. Clarke, and A. Cienki, "Sensegiving in entrepreneurial contexts: The use of metaphors in speech and gesture to gain and sustain support for novel ventures," *Int. Small Bus. J.*, vol. 30, no. 3, pp. 213–241, 2012.
[54] A. W. Brooks, L. Huang, S. W. Kearney, and F. E. Murray, "Investors prefer entrepreneurial ventures pitched by attractive men," *Proc. Nat. Acad. Sci. United States Amer.*, vol. 111, no. 12, pp. 4427–4431, 2014.
[55] C. Cameron and J. Dickfos, "Lights, camera, action! Video technology and students' perceptions of oral communication in accounting education," *Accounting Educ.*, vol. 23, no. 2, pp. 135–154, 2014.
[56] K. Lucas, S. A. Kerrick, J. Haugen, and C. J. Crider, "Communicating entrepreneurial passion: Personal passion vs. perceived passion in venture pitches," *IEEE Trans. Prof. Commun.*, vol. 59, no. 4, pp. 363–378, Dec. 2016.
[57] M. A. McCollough, B. Devezer, and G. Tanner, "An alternative format for the elevator pitch," *Int. J. Entrepreneurship Innov.*, vol. 17, no. 1, pp. 55–64, 2016.
[58] M. P. Ciuchta, C. Letwin, R. Stevenson, S. McMahon, and M. N. Huvaj, "Betting on the coachable entrepreneur: Signaling and social exchange in entrepreneurial pitches," *Entrepreneurship Theory Pract.*, vol. 42, no. 6, pp. 860–885, 2017.
[59] J. S. Clarke, J. P. Cornelissen, and M. P. Healey, "Actions speak louder than words: How figurative language and gesturing in entrepreneurial pitches influences investment judgments," *Acad. Manage. J.*, vol. 62, no. 2, pp. 335–360, 2019.
[60] J. M. Jachimowicz, C. To, S. Agasi, S. Côté, and A. D. Galinsky, "The gravitational pull of expressing passion: When and how expressing passion elicits status conferral and support from others," *Organiz. Behav. Human Decis. Processes*, vol. 153, pp. 41–62, 2019.
[61] W. R. Berdine, "Planning and structuring a business speech," *Bus. Commun. Quart.*, vol. 36, no. 4, pp. 28–32, 1973.
[62] B. A. Andeweg, J. C. de Jong, and H. Hoeken, "May I have your attention?: Exordial techniques in informative oral presentations," *Tech. Commun. Quart.*, vol. 7, no. 3, pp. 271–284, 1998.
[63] D. Kirsch, B. Goldfarb, and A. Gera, "Form or substance: The role of business plans in venture capital decision making," *Strategic Manage. J.*, vol. 30, no. 5, pp. 487–515, 2009.
[64] A. B. Thro, "CEOs' hybrid speeches: Business communication staples," *J. Bus. Commun.*, vol. 46, no. 3, pp. 335–361, 2009.
[65] K. Klyver, E. Hunter, and T. Watne, "Entrepreneurial ties and innovativeness in the start-up decision," *Int. J. Entrepreneurship Innov.*, vol. 13, no. 3, pp. 153–163, 2012.
[66] J. M. Pollack, M. W. Rutherford, and B. G. Nagy, "Preparedness and cognitive legitimacy as antecedents of new venture funding in televised business pitches," *Entrepreneurship Theory Pract.*, vol. 36, no. 5, pp. 915–939, 2012
[67] C. Spinuzzi *et al.*, "Remaking the pitch: Reuse strategies in entrepreneurs' pitch decks," *IEEE Trans. Prof. Commun.*, vol. 58, no. 1, pp. 45–68, Mar. 2015.
[68] P. Daly and D. Davy, "Crafting the investor pitch using insights from rhetoric and linguistics," in *The Ins and Outs of Business and Professional Discourse Research: Communicating in Professions and Organizations.* G. M. Alessi and G. Jacobs, Eds. London, UK: Palgrave Macmillan, 2016.
[69] P. Daly and D. Davy, "Structural, linguistic and rhetorical features of the entrepreneurial pitch: Lessons from Dragons' Den," *J. Manage. Develop.*, vol. 35, no. 1, pp. 120–132, 2016.
[70] T. Ladd, "Does the business model canvas drive venture success?" *J. Res. Market. Entrepreneurship*, vol. 20, no. 1, pp. 57–69, 2018.
[71] J. C. Sort and C. Nielsen, "Using the business model canvas to improve investment processes," *J. Res. Market. Entrepreneurship*, vol. 20, no. 1, pp. 10–33, 2018.
[72] H. Gafni, D. Marom, and O. Sade, "Are the life and death of an early-stage venture indeed in the power of the tongue? Lessons from online crowdfunding pitches," *Strategic Entrepreneurship J.*, vol. 13, no. 1, pp. 3–23, 2019.
[73] S. Shane, W. Drover, D. Clingingsmith, and M. Cerf, "Founder passion, neural engagement and informal investor interest in startup pitches: An fMRI study," *J. Bus. Venturing*, vol. 35, no. 4, 2019, Art. no. 105949.

[74] J. A. Ward, "Assessing creativity in Shark Tank entrepreneurial pitch meetings: Extending dual-process judgment theory," *Acad. Entrepreneurship J.*, vol. 21, no. 2, pp. 119–132, 2015.
[75] B. Smith and A. Viceisza, "Bite me! ABC's Shark Tank as a path to entrepreneurship," *Small Bus. Econ.*, vol. 50, no. 3, pp. 463–479, 2018.
[76] J.-S. Fernández-Vázquez and R.-C. Álvarez-Delgado, "The interaction between rational arguments and emotional appeals in the entrepreneurial pitch," *Int. J. Entrepreneurial Behav. Res.*, vol. 26, no. 3, pp. 503–520, 2019.
[77] J.-S. Fernández-Vázquez and R.-C. Álvarez-Delgado, "Persuasive strategies in the SME entrepreneurial pitch: Functional and discursive considerations," *Econ. Res.*, vol. 33, no. 1, pp. 2342–2359, 2020.
[78] R. van Werven, O. Bouwmeester, and J. P. Cornelissen, "Pitching a business idea to investors: How new venture founders use micro-level rhetoric to achieve narrative plausibility and resonance," *Int. Small Bus. J.*, vol. 37, no. 3, pp. 193–214, 2019.
[79] R. E. Burns, "The vital importance of the elevator speech," *Bridgewater Rev.*, vol. 35, no. 2, pp. 33–34, 2016.
[80] T. H. Byers, R. C. Dorf, and A. J. Nelson, *Technology Ventures: From Idea to Enterprise*. New York, NY, USA: McGraw-Hill, 2010.
[81] G. Elia, A. Margherita, and C. Petti, "An operational model to develop technology entrepreneurship ego-system," *Int. J. Innov. Technol. Manage.*, vol. 13, no. 5, pp. 1–23, 2016.
[82] H. Jo and J. Lee, "The relationship between an entrepreneur's background and performance in a new venture," *Technovation*, vol. 16, no. 4, pp. 161–211, 1996.
[83] E. Goffman, *The Presentation of Self in Everyday Life*. New York, NY, USA: Doubleday Ancho, 1959.
[84] H. E. Aldrich and C. M. Fiol, "Fools rush in? The institutional context of industry creation," *Acad. Manage. Rev.*, vol. 19, no. 4, pp. 645–670, 1994.
[85] J. A. Katz and R. P. Green, *Entrepreneurial Small Business*. New York, NY, USA: McGraw-Hill/Irwin, 2008.
[86] D. B. Audretsch, "From the entrepreneurial university to the university for the entrepreneurial society," *J. Technol. Transfer*, vol. 39, no. 3, pp. 313–321, 2014.
[87] J. Heinonen and U. Hytti, "Back to basics: The role of teaching in developing the entrepreneurial university," *Int. J. Entrepreneurship Innov.*, vol. 11, no. 4, pp. 283–292, 2010.
[88] S. C. de Janasz and M. L. Forret, "Learning the art of networking: A critical skill for enhancing social capital and career success," *J. Manage. Educ.*, vol. 32, no. 5, pp. 629–650, 2008.
[89] T. Bager, "The camp model for entrepreneurship teaching," *Int. Entrepreneurship Manage. J.*, vol. 7, no. 2, pp. 279–296, 2011.
[90] D. Ballantyne, P. Frow, R. J. Varey, and A. Payne, "Value propositions as communication practice: Taking a wider view," *Ind. Market. Manage.*, vol. 40, no. 2, pp. 202–210, 2011.
[91] H. Berglund, D. Dimov, and K. Wennberg, "Beyond bridging rigor and relevance: The three-body problem in entrepreneurship," *J. Bus. Venturing Insights*, vol. 9, no. 5, pp. 87–91, 2018.

Alessandro Margherita received the Ph.D. degree in Management Engineering from the University of Salento, Lecce, Italy, in 2007. He is currently a Senior Researcher with the Department of Engineering for Innovation, University of Salento. He was a Research Affiliate and Visiting Scholar with MIT Sloan and the Peking University, and he is also a Research Partner with the Federal University of Santa Catarina, Florianópolis, Brazil. He has authored several management books and published in journals, such as the IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, *Technology Forecasting and Social Change*, *Expert Systems With Applications*, and *Human Resource Management Review*. His research interests include digital transformation and technology entrepreneurship.

David Verrill received the bachelor's degree in Biology from Bowdoin College, Brunswick, ME, USA, in 1983, and the M.Sc. degree in Management from MIT Sloan, Cambridge, MA, USA, in 1987. He is currently the Executive Director with the MIT Initiative on the Digital Economy, Boston, MA, USA. He has worked and consulted with MIT for more than 30 years. Prior to his work there, he was a Research Scientist with the Center for Blood Research and Xerox Adaptive Products, where he was the Manager of the International Sales and Business Development. He has helped start two companies, including Winchester International Group and the Hub Investment Group, which have invested in more than 50 early-stage companies in the New England region. He sits on the boards of several startups, is the Chairman Emeritus of the Angel Capital Association, and is the recipient of the Hans Severiens Award.

Research Article

The Critical Communication Challenges Between Geographically Distributed Agile Development Teams: Empirical Findings

—Yehia Alzoubi and Asif Gill

***Abstract*—Background:** *Although a number of empirical studies have investigated communication challenges during recent years, we still need to discover the most critical challenges that face communication when agile development is geographically distributed. We also need to discover how successful geographically distributed agile development (GDAD) organizations deal with these challenges.* **Literature review:** *Most previous studies reported that the critical challenges facing GDAD communication can be categorized into five themes: differences in cultures, different time zones, different spoken languages, different personal skills, and the efficiency and effectiveness of communication tools used.* **Research questions:** *1. What are the challenges of communication between GDAD teams? 2. How can the impact of GDAD communication challenges be mitigated?* **Methodology:** *Data were collected by interviewing 12 members of a three-team organization using distributed agile development. These teams are distributed over three countries; the main team located in Australia, the developers' team located in China, and the testers' team located in India. A thematic analysis technique was used to identify communication challenges and practices used to mitigate the effect of these challenges.* **Results:** *Our findings reveal that the five challenges are still critical to GDAD. Moreover, we report a new critical challenge of communication in GDAD, the insufficient documentation provided by distributed teams and members. In addition, we recommend several practices to mitigate the impact of these challenges.* **Conclusions:** *Communication among distributed agile development teams still faces several critical challenges, and the solutions to these challenges provided in recent years have not been sufficient. This fact prompts the need for more research on how the impact of these challenges can be lessened.*

Index Terms*—Challenge, communication, geographically distributed agile development (GDAD), recommendation.*

Many agile software development (ASD) firms have adopted geographically distributed agile development (GDAD) because of the success of colocated agile development and the many benefits of the distributed environment, such as cost, availability of skilled developers, and proximity to markets [1], [2]. However, because of the challenges facing GDAD, the failure rate of these projects is high [1], [3], [4]. Communication is the main problem that faces GDAD [5]–[7], resulting in a lack of coordination and collaboration among distributed teams, and understanding of customer requirements [5]. The concept underlying ASD is its people- and communication-oriented approach [1], [8], [9]. Unless distributed teams and team members achieve high communication efficiency and effectiveness, the success of the GDAD project is uncertain [10], [11].

Several challenges of GDAD communication have been identified in the literature, such as less face-to-face informal communication, time differences, language barriers, different cultures, and different locations [12]–[15]. Although GDAD communication has been a subject of interest among researchers, many issues are still unaddressed [6], [13], [16]. The communication issues being faced by GDAD still present a big challenge to most GDAD projects, and as yet, no convenient solutions addressing this issue have been provided [13], [17], [18].

Several studies have tried to empirically investigate how these challenges to distributed agile teams may be solved in a successful large agile development organization [13], [14], [19], [20]. However, most prior studies have examined only particular challenges such as communication with customers, language barriers, time differences, or

Manuscript received November 8, 2020; revised April 8, 2021; accepted April 9, 2021. Date of publication October 21, 2021; date of current version November 24, 2021. *(Corresponding author: Yehia Alzoubi.)*
Yehia Alzoubi is with the College of Business, American University of the Middle East, Egaila 15453, Kuwait (email: yehia.alzoubi@aum.edu.kw).
Asif Gill is with DigiSAS Lab, School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia (email: asif.gill@uts.edu.au).

IEEE 10.1109/TPC.2021.3110396

Practitioner Takeaway

- The findings reveal six critical communication challenges: culture differences, time differences, language barriers, tools and technology, personal skills, and documentation.
- The new challenge discussed in this paper (insufficient documentation) has not been reported in the literature before.
- The paper outlines strategies for reducing the negative effects of these challenges: training on cultural nuances, time overlapping between dispersed teams, learning and training on business language, keeping communication tools up to date, and providing adequate documentation.

individual skills [12], [21]. Because there is still a need for more GDAD research, this study empirically investigates the most critical challenges and presents recommendations to mitigate the effect of these challenges. More precisely, this study aims to answer the following research questions.

RQ1. What are the challenges of communication between GDAD teams?
RQ2. How can the impact of GDAD communication challenges be mitigated?

The first question focuses on identifying the critical challenges that decrease or hinder communication between teams in large successful GDAD organizations. The second question identifies how GDAD organizations overcome or decrease the effects of these challenges and enhance the efficiency and effectiveness of communication between their distributed teams. In pursuing these research questions, we used an in-depth exploratory case study.

Our case organization was a large software development company that includes 39 members, with the main location in Sydney (main team) and the other two teams located in China and India. Over a two-month period, we collected data through semistructured interviews and observation of the main team. We interviewed 12 key participants distributed over the three locations. We used a thematic analysis technique to identify the GDAD communication challenges and recommendations to mitigate the effects of these challenges.

This article enriches the understanding of GDAD communication challenges by contributing to the community of researchers and practitioners by providing a close in-depth look at these challenges. We uncover a particular challenge, insufficient documentation, which, to our knowledge, has never been mentioned as a challenge facing GDAD communication in previous literature. This may represent an important contribution because if recognized as a challenge, it may help practitioners in large GDAD projects and open new avenues for researchers. Furthermore, this article prioritizes these challenges and reveals that cultural differences among distributed teams are the most critical communication challenge. Moreover, this study discusses the techniques used by a successful GDAD firm to mitigate the impact of these challenges.

The rest of this article is organized as follows. First, we review the research background and then discuss the research design. Next, we present the findings of this study, followed by its implications for practice and research and the study's limitations. Finally, we discuss our conclusions.

Literature Review

GDAD Communication Communication generally refers to exchanging information and knowledge that aims at achieving a common understanding between the sender and receiver [22], [23]. Communication is a crucial part of all software development [4], [24]–[26]. ASD emphasizes and counts on direct communication among team members and customers. Direct and informal communication is beneficial because it increases the speed of interactions and maintains strong relationships and trust levels among members [27]. Some agile approaches depend on a high level of informal direct communication [28]. Approaches such as eXtreme Programming (XP) and Scrum were reported as beneficial because they enhance direct communication and promote less documentation among team members [29].

Therefore, ASD can facilitate both formal and informal communication [30]. Scrum was reported as an enabler of higher quality and frequency of communication among team members because it provides a framework that reminds and promotes members to interact regularly and closely [28], [29].

Communication is even more critical in GDAD because the direct "face-to-face" and informal meetings among distributed teams are less frequent. Because GDAD refers to distributed (virtual) or noncolocated teams [31], teams can be distributed in more than one location either inside the same time zone (e.g., in the same country) or over different time zones (e.g., in different countries) [24], [32]. GDAD is different from colocated agile development in several characteristics, including temporal distance (i.e., different time zones or time-shifting work), geographical distance (i.e., different physical locations), and sociocultural distance (i.e., the cultural differences between distributed teams or members) [22], [33], [34]. These characteristics relate directly and negatively to the effectiveness and efficiency of GDAD communication, which results in difficulties in scheduling and delays in feedback [35]. This may lead to misunderstanding of tasks and project requirements, effort duplication, and running behind schedule or above the cost [10], [36]. Accordingly, managing communication is essential to ensure successful projects in GDAD [37]. Moreover, unlike colocated ASD, pure informal communication may be difficult or even impossible to achieve in GDAD [38]. That is, some level of formal communication is required, especially when the number of stakeholders increases in large projects, and that fact can, in fact, hinder communication and coordination [38]. Thus, direct communication may not always be the best choice and may become an inhibitor [39].

GDAD communication depends more on indirect technology-mediated channels rather than direct communication channels. Therefore, the successful delivery of the GDAD project will not be achieved without the successful utilization of communication technologies [40]. Practically, it is important to choose the right communication channel because these technologies may create another obstacle in communication effectiveness [28]. Using communication technologies instead of direct communication may decrease the trust level among distributed members because it eliminates the communication cues that enhance trust when direct communication occurs [41]. Moreover, communication using these technologies may face a problem when the size of the GDAD team increases because choosing the right technology that guarantees participation and awareness will be problematic [42].

Related Work Several studies have empirically investigated GDAD communication challenges (e.g., [13], [14], [19], [20]). Ammad et al. [13] ran a systematic review to identify the GDAD communication issues. They identified eight categories of challenges: sociotemporal distance (time difference), geographical distance (i.e., distribution over different locations), organizational and architectural issues (i.e., the misunderstanding or unnecessary flow of communication due to the definition of the system and software structures), sociocultural distance, team members' attitudes, team issues, technical issues, and customer issues. These challenges were further tested by running a survey among GDAD developers. The findings suggested that all these challenges had a significant effect on GDAD projects, except for the technical issues.

Alzoubi et al. [12] developed the GDAD Communication Framework based on a systematic literature review of the empirical studies that investigated GDAD communication challenges. This framework identifies 17 GDAD communication challenges and organizes them into six major categories.

1. Distance differences, including different time zones and different geographical areas
2. Team configuration, including team size, number of teams, and coordination among teams
3. Project characteristics, including project domain and project architecture
4. Customer communication, including involvement of customer and customer representative
5. Organizational factors, including project management processes, communication tools, communication infrastructure, and organizational culture
6. Human factors, including language, national culture, trust, and personal practices

Martini et al. [19] investigated GDAD communication challenges among agile developers in three GDAD firms using surveys. The authors studied five challenges: technology (programming languages and tools used by GDAD teams), architecture, organization (coordination, task allocation, and supervision), process (delivering tasks), and people factor (including personal or

group attitudes, mindset, and knowledge). All of these factors except for technology were found to be significant and to negatively affect GDAD communication. The authors also suggested some solutions to mitigate the impact of these factors. For the architecture factor, the authors suggested that the high-level system requirements should be agreed on at the beginning of the project and reference architecture should be deployed. For the process factor, the authors suggested providing and coordinating frequent interfaces for strategic aspects among distributed team members. For the team distribution factor, the authors suggested exchanging visits and promoting the use of social media. For the organization factor, avoiding organizational dependencies caused by the architecture as well as relocating distributed team members was suggested to create relationships and increase knowledge sharing among members of different teams.

Dorairaj et al. [20] investigated GDAD communication challenges by interviewing two teams: one in the USA and one in India. Four challenges were identified: time difference, language barriers, teamwork tasks, and communication tools. For the time difference challenge, the authors suggested minimizing time differences among team members. For language barriers, the following strategies were suggested: informing team members about the meeting topics in advance and speaking slowly during meetings. Enhancing trust and promoting formal communication were suggested to decrease the effect of the challenges facing teamwork tasks. Finally, for the communication tools issue, the authors suggested tailoring and systematically deploying communication tools to fit the GDAD purpose.

Layman et al. [14] investigated GDAD communication challenges through interviewing developers in two XP teams distributed in the US and the Czech Republic. They studied the effect of the time difference, geographical distance, language barriers, and communication tools. The authors provided some recommendations, in general, to mitigate the effect of these communication challenges among distributed teams, such as using asynchronous tools, encouraging team members to work closely with project management on a daily basis, and recording and tracking daily progress using the available project management tools.

This study investigates a large GDAD organization that accomplished several successful projects, as described by participants. This study adds to the body of literature by identifying the issues facing communication between its distributed teams. It also identifies and investigates the insights of how such a successful GDAD firm manages to mitigate the impact of these issues. This is of high importance to GDAD firms since many GDAD projects have been unsuccessful because of communication issues between distributed teams [12].

Research Design

In this research study, we used a single case-study method because it provides an in-depth understanding of a real-life phenomenon and helps discover hidden patterns that cannot be unveiled using other methods such as that described by Yin [43]. According to Yin [12], a case study should include study questions, propositions, logical links between the data and the propositions, and a unit of analysis. The questions asked are shown in Section Data Analysis and Coding, the proposition is that GDAD communication will be decreased by some challenges, the logical linking of this article to literature is that the identified challenges' impact can be mitigated by some practices, and the unit of analysis is the GDAD team member.

Research Context The data were collected from a large company that specializes in delivering new applications and whose main team (18 members) is located in Sydney, Australia. In this article, we will refer to this company as SUNC to preserve anonymity. SUNC has been using GDAD for more than seven years and has adopted the Scrum method of agile development. Scrum is the most preferable agile approach in GDAD since it provides a level of management and control [12]. SUNC employed distributed teams of developers and testers on each project.

This article studied only the pensioner banking system (PBS), which was under development during data collection. SUNC started this project in late 2016. SUNC employed two distributed teams: the China team, with 16 members (including solution architecture, Scrum Master, and developers), and the India team, with 5 members (including infrastructure architect, test leader, and three testers). The Sydney team, with the most experienced members, according to interviewees in Sydney, included the program manager, integration manager, enterprise architect, business architect, delivery manager, iteration manager, product owner, Scrum master leader, analysts, reviewers, and developers. The majority of the work was done

TABLE I
INTERVIEWEE DEMOGRAPHICS

Participant	Role	Location	Experience (Years)	Interview Method
A	Program manager	Sydney	10	Face-to-face
B	Enterprise architect	Sydney	9	Face-to-face
C	Product Owner	Sydney	7	Face-to-face
D	Scrum Master leader	Sydney	8	Face-to-face
E	Integration manager	Sydney	11	Face-to-face
F	Tester Leader	Sydney	6	Face-to-face
G	Iteration manager	Sydney	8	Face-to-face
H	Developer	Sydney	4	Face-to-face
I	Solution architect	India	7	Skype
J	Tester	India	6	Skype
K	Scrum Master	China	4	Skype
L	Technology Architect	China	5	Skype

at the Sydney location, such as collecting requirements, assigning tasks, integration, and configuration. The Sydney team described PBS as a big project requiring about two years to complete. SUNC's policy was that all new joiners should report to the Sydney location for one-month's training to familiarize them with the system and the technical language used. Table I provides the demographics of all interviewees in this study. To preserve anonymity, all participants were assigned alphabetical codes.

PBS Process and Team Roles All teams followed the core Scrum practices with three-week Sprints. Every four Sprints (12 weeks), the three teams released a product. All teams were required to attend the Sprint planning meeting at the beginning of each sprint. The meeting was usually conducted at 2:00 P.M. Sydney time to overcome the difference in time between distributed teams (the time difference between Sydney and the China team was 3 hours and between Sydney and the India team was 5 hours). Usually, sprint information was shared in the form of artifacts (sprint goals, sketches of the project plan, the finished tasks, and so on) on the walls at the Sydney location. JIRA was used as a tool to share the stories of the workflow among the three distributed teams. JIRA is a tool developed by Atlassian Company that supports Scrum practices and provides time-tracking capabilities and real-time performance reports in GDAD [44]. SUNC also has an online repository to share the code and refinements between teams.

All planning, instructions preparation, and task assignments for all teams were undertaken by the Sydney leaders. The product owner at the Sydney location assigned tasks to the China team, which was responsible for functional and technical components. The product owner and Scrum master leader at the Sydney location synchronized to write the user stories and to group them before sending them to the China team. Although the requirement discovery was continuously updated, it was most intensive for the two weeks before the start of the new release. Business requirements were identified and reviewed by different members, including the program manager, enterprise architect, delivery manager, and customer. Usually, the China team would take about two weeks to finish a task before sending it for testing by the India team, which provided refinement of the features, data, or user interfaces. Then, the task was sent to the Sydney team for review with the customer before integrating the new release with the other parts of the project. The program manager communicated with the customer either face-to-face or using electronic tools. The delivery manager was responsible for evaluating risks related to the current sprint, while the iteration manager was responsible for managing most team meetings.

Data Analysis and Coding Thematic analysis was used to analyze the data collected [45]. The

thematic analysis technique enables identifying, analyzing, and reporting themes within the data [46]. All of the important comments across all participants were identified and compared, and comments were contrasted and compared to draw the whole picture about shared patterns and key insights [45]. To systematically analyze and code the collected data, we followed the guidelines provided by Miles and Huberman [45], which include four stages: data collection, data reduction, data display, and conclusion drawing.

Data Collection: We sent emails to all interviewees and obtained their confirmation to conduct the interview. Most interviews were done face-to-face in the main team (eight interviews at the Sydney location). The remaining interviews were conducted using Skype. All interviews were conducted in English. We obtained participant consent with a form and followed an interview protocol in all interviews. Data were collected mainly through semistructured interviews, which allowed interviewees to express their views in their own way and report more insights [43]. Additional data were collected through the notes taken during a two-month observation of the main team site in Sydney, including sprint planning and regular colocated meetings as well as notes from some artifacts used, such as charts and sketches. Each interview lasted 50–70 minutes. We started each interview by asking general questions such as the interviewee's role or responsibility in the PBS project, and then we asked the interviewee the following predefined questions. The purpose of these questions was to identify the communication challenges and the techniques used to mitigate these challenges among distributed teams in the PBS project.

- What challenges/difficulties do you face when you communicate with distributed teams/team members?
- How can you solve or mitigate the effect of these challenges?
- What techniques/practices/tips could SUNC use to mitigate the effect of these challenges?
- What other considerations/tools/techniques/strategies would you like to be included in SUNC when communicating with distributed teams?

Data Reduction: After collection, data should go through a preparation and familiarization process that makes it ready for analysis. Consequently, all audio records, as well as the notes collected during team observations in the Sydney location, were transcribed into a Word document. Moreover, the handwritten notes and information that were taken by researchers during the interviews were used as additional information and guidance for the analysis. Furthermore, the researchers took a close look at the data to become familiar with its richness, depth, and diversity. Familiarization was also achieved throughout the process of reviewing, organizing, and transcribing the data.

Data reduction occurred continually throughout the analysis process [45]. To transfer the data into a usable form, preliminary coding was essential. This coding included a preparatory data reduction in which we discarded personally identifiable information and the appreciation sections of each interview. Then, we identified the main themes. Here, the data were examined to identify key issues, concepts, and themes that appeared related to the research questions. This process was done iteratively within and across case data. The coding scheme indexed the text related to the presence of GDAD communication and communication challenges, and communication-related mechanisms and tools. Systematic coding allows researchers to see the logical link between the theoretical research model and the codes and provides a means to avoid bias and validate interpretations [45]. The coded data were then grouped into related categories and linked for thematic examination.

Data Display: Data display organizes, compresses, and assembles the data into a more readable format than the original format, where data are scattered everywhere [47]. Therefore, data display was done at all stages of the analysis. Data display is also concerned about the stage of analysis that has been achieved and the preparation for the next stage of analysis [47]. Each set of interview codes was examined to identify the elements that were primary to the category. These codes were then mapped into tables and diagrams for comparison with the elements in the reference model.

Conclusion Drawing: The final stage is drawing conclusions and involves synthesizing tables and charts against the predefined themes/categories. Data coding, categorization, and identification of major themes were mapped to the challenges identified on the GDAD Communication Framework developed by Alzoubi et al. [12], as discussed in the Literature Review section. However, because the focus of this study is communication among distributed teams and team members, customer communication-related challenges were not included. This framework was

used because it aligns very well with the focus of this study. Moreover, this framework is able to accommodate the most important aspects of GDAD communication challenges, recommendations that may mitigate the impact of these challenges, and the impact of the challenges and solutions on the GDAD communication process. The aim of this final stage is to create a coherent meaningful form of the collected data. This stage occurred concurrently with other stages during the analysis [45].

Findings

The purpose of the communication process in agile development is to communicate project requirements, define goals for the task definition, and deliver the solutions or projects. Tasks were written and described on the wall of an open office space in Sydney using the task board (based on the customers' stories). The onsite team and distributed teams could see the tasks either by attending meetings in person or by viewing videoconferences. When each iteration began, the task board was used as a communication tool, and information sharing between distributed teams increased the visibility of the common project goals. This practice allowed continuous tracking of the product backlog, work tasks, and work processes. Moreover, the sprint backlog artifacts, such as the Excel sheet that included project tasks and estimates for the next iteration, represented a valuable practice for daily project status evaluation in the iteration retrospective meeting. As Participant A described, "it is easier to discuss tasks and see the progress of my team using the storyboard." The communication challenges, as well as some tips and techniques to overcome those challenges, revealed through the analysis of interviews and team observation, are discussed in the following sections. A summary of the findings is shown in Fig. 1.

Communication Challenges In-depth analysis of the data collected through interviews and observation by the Sydney team identified six challenges to communication among the three teams. The major sources of communication issues were cultural differences (reported by eight interviewees), time difference (reported by five interviewees), language barrier (reported by five interviewees), communication technologies and tools (reported by three interviewees), personal skills (reported by two interviewees), and insufficient documentation (reported by two interviewees).

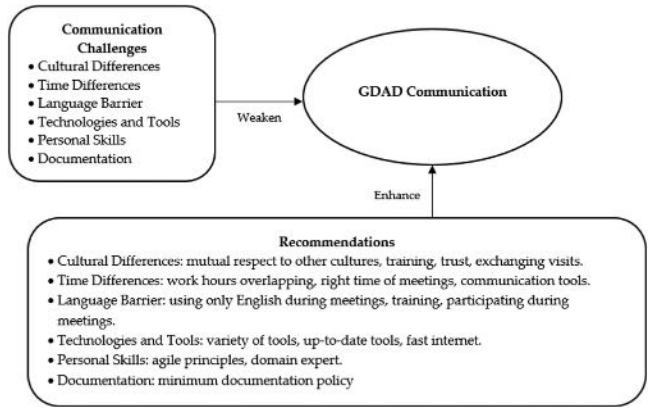


Fig. 1. GDAD communication challenges and recommendations/practices.

Cultural Differences (National and Organizational Cultures): National culture refers to "the collective programming of the mind that distinguishes the members of one group or category of people from another" [12, p. 31]. Organizational culture refers to "the values, attitudes, and behaviors that represent an organization's working environment, vision, and subjectivity" [12, p. 30]. This challenge was identified by several studies (e.g., [12], [13] and was the most reported by the interviewees. Most interviewees agreed to a great extent that national cultural difference was the biggest problem in communication among the three teams. This problem may have resulted from the fact that these teams were globally distributed over three countries (including multicultural teams). Participant B said that

> the biggest problem we face is cultural differences. While the team in India is very hierarchical and does not ask many questions during and after meeting, the team members in China are very reluctant to ask questions during the meetings; however, they will come with many questions after the meeting.

The national cultural differences required special skills of the Sydney team leaders when communicating with different cultural teams. This fact, of course, added more value to their expertise. Participant C explained that

> both teams in India and China should be dealt with differently because of the difference in culture. You need to be very careful of the way you address some issues so your message is not understood personally. Also, what and how do you report things … in our communication repository and tool.

Another issue related to culture was the organizational culture that may negatively impact communication and coordination among distributed teams. This issue was related to how each distributed team did work in the local context. According to Participant E, "the other important issue of communication with off-shore teams is the organization culture."

Time Differences: This challenge was identified in several studies (e.g., [13], [14], [20]). This challenge was the second most reported by the interviewees. As mentioned before, the greatest time difference between teams was 6 hours. So they had to arrange each team's work hours to enable all teams to participate in the meetings. The Sydney team started working at 10:00 A.M. and finished at 6:00 P.M., Sydney time. The China team started at 11:00 A.M. and finished at 7:00 P.M., Sydney time. Finally, the India team started at 12:00 P.M. and finished at 8:00 P.M., Sydney time. These arrangements provided a 6-hour overlap of working hours.

Nevertheless, the time difference was assumed to be the main challenge by some interviewees. This challenge resulted in less synchronized communication and decreased the frequency of communication in general among teams. According to Participant E, "the main issue with communication with off-shore teams was the time difference, which makes it very hard to make communication any time."

Language Barrier: The language barrier refers to the different native spoken languages among distributed teams [12]. This challenge has been identified by several studies (e.g., [12], [14], [20]). This challenge was the third most reported by the interviewees. Since not all members were native English speakers, communication among teams was described as poor. This was true even if developers were skillful in the business lingua franca (English). Team members in China and India preferred to speak their local languages within their teams. They spoke English only when they had meetings with other teams (e.g., a sprint planning meeting). Participant E reported that "The other important issue with communicating with off-shore teams is the 'language barrier' although everyone [who] works for our company should speak English."

Moreover, the language issue was still noticeable even when multicommunication channels were used among distributed teams. As Participant F stated, "There is a public forum and also private channels for communications; these are effective; however, most of the problems stemmed from the language barrier."

Technologies and Tools: The fourth issue that affected communication among distributed teams in PBS was using the communication tools themselves. This challenge has been identified in previous studies (e.g., [13], [20]). Although most practitioners and scholars recommend using these tools as a substitute for face-to-face communication, Participant G clearly stated that

> Nothing is going to be like face-to-face communication. However, you may reach that, but that will take a very long time. The communication tools should be used properly to give the maximum benefits.

Communication tools may negatively impact team communication if the tools are not tailored and systematically deployed to fit the purpose of GDAD. This situation will result in less efficient and effective GDAD communication. Participant H reported that "It is always about [using] the right technology to support conversation."

Moreover, these tools were used to record and track the daily development progress. As a result, some members were not always ready or happy to do synchronous communication. Participant D reported that

> One of the issues of communicating with off-shore teams would be the technology (tools), I think. Every member should be connected to "TV Link" and keep the web-cam on during the session time to share their screens with other members. The … webcam and recording the meetings may not really be welcomed by some members.

Personal Skills: Personal skills refer to the difference in developing skills and experience among distributed members. This challenge was identified in previous studies (e.g., [12], [19]). In a large GDAD team, not all members have the same skills, a problem that decreases or hinders communication with external distributed members [12]. Differences in personal skills require an increase in the frequency of communication and the time required to discuss or share knowledge. As a result, communication efficiency will be lower. Alaiad et al. [36] emphasize the role of personal skills on the communication and performance outcomes of virtual teams. They reported that most prior studies did not look at the individual

characteristics of the team members, such as skills [36].

Another challenge was identified as another challenge for communication in PBS. Participant E reported, "One of the issues with off-shore communication would be the difference in team members' experience." Most interviewees from the Sydney team thought that personal skills decreased with distance such that the most skilled members were those who worked on that team, followed by the those who were located in the distributed team inside the same country, and then by members in off-shore locations (i.e., globally distributed in different countries).

Therefore, the onsite team had to help distributed teams and show them how to do some tasks if they did not have the necessary skills. Participant C claimed, "The main team represents a hub to all distributed teams in order to mitigate the skills differences." Accordingly, the Sydney team should have the main resources to manage and help other teams do their tasks. Participant G claimed that the

> Sydney team has the majority of the resources since it works as a hub to all other distributed teams. All teams can watch and see what we do here in Sydney. This needs more effort from members in Sydney, but it enables all distributed teams to capture everything that happened here.

Documentation: This challenge refers to insufficient documentation written and used by various teams. In colocated ASD, due to high efficiency and effectiveness of communication among team members, less documentation leads to lower overhead [48]. However, brief documentation that is just enough to serve its purpose is very important [48]. This is even more essential in GDAD, where the chance of face-to-face meetings is less, team size is large, and multicultural and multilingual members' involvement is high. Therefore, minimum documentation is required to keep all members on the same page [42]. Accordingly, the least documentation principle of ASD may not be totally applicable in GDAD, and without the required level of documentation, GDAD communication efficiency and effectiveness will be lower [38], [42].

To our knowledge, this challenge has not been identified by previous studies. Less documentation of the main artifacts was assumed to be another challenge for communication by some interviewees. That is, face-to-face communication, in the colocated context, saves time and effort and reduces documentation, and that fact, in turn, increases the benefits for business and enhances customer satisfaction. However, in the GDAD environment and due to lack of face-to-face communication, documentation of main artifacts and codes is important. Participant B reported that "Distributed developers do not like to document what they do, which may represent another challenge to future communication if we need to do some changes or if we have new joiners." In fact, without these artifacts, distributed members might not understand what they were supposed to do. Participant D said,

> Definitely, otherwise, how will they handover? You need at least minimal documentation, but not too much, and [it] should be easily searchable, understandable, and accessible by all stakeholders.

Recommendations/Practices to Enhance GDAD Communication Based on the in-depth analysis of the data collected through interviews and observation of the Sydney team, a number of techniques were identified to mitigate the GDAD communication challenges.

Cultural Differences: Some techniques were recommended by interviewees to overcome the problem of cultural differences. These techniques depended first on the experience (i.e., communication skills) of the team members, on the executives' strategies, and on the mutual respect for other members' cultures when establishing communication. Participant D reported,

> You need to make sure that the message has been received as it was intended to be. If you rush and do not take your time or you do not ask the question in the right way, you will have a misunderstanding of your message. You need to take away the barriers between you and them, so talk. So you need to ask different types of questions, such as "did you get it?" "Say yes if you get it." "Say no if no," and so on.

Moreover, the team leader must ensure that all members share and communicate with others. Participant D added, "ask them and let them tell what they did yesterday, what they are doing today, and what they will do tomorrow. If they have a problem, ask them what I can do to resolve the problem."

Second, other interviewees recommended cultural training for team members, especially in the

beginning when joining the team. This was made clear by Participant C, who reported, "Yes, we do cultural training for all new members when they first [join] the team." Another recommended technique was exchanging cultural knowledge among distributed teams about the differences between cultures. This was done by asking team members to tell other members about their own cultures. Participant I commented,

> We have the "buy-in," which is a cultural thing to make different cultural distributed teams tell us how things should work. You need to be flexible with other cultural teams. Agile is all about flexibility.

The third technique was by exchanging visits, which helped in developing a better understanding of other distributed teams. Participant L reported that

> The new team members should go to Sydney and spend some time at the beginning of their joining our teams. Also, the leaders from the Sydney team [should] visit different distributed teams regularly.

Finally, trust and respect for other cultures were reported as critical values that should be practiced to overcome cultural issues. Moreover, it was reported that the new team members should be introduced to everyone to encourage them to adopt the common team spirit. Participant E explained, "Our Company has a great culture such that everyone shows high respect to other teams or members wherever they are."

Time Differences: To decrease the impact of the time difference, SUNC teams usually adjusted their working hours to reach the maximum amount of overlap. Participant J reported, "We do overlap between teams." Even so, sometimes it was hard to get all distributed teams in one meeting at the same time. As reported by Participant K, they "… make sure the timing of those conversations is optimum." And Participant D added, "sometimes we just make sure that the majority attend at a time. However, teams' members can ask questions anytime." However, as reported by Participant A,

> it is always a challenge to find the right time for overlap since it is not sustainable to ask team members to work from 11:00 A.M. to 10:00 P.M., for example, to get all distributed members in the one-time meeting.

Another technique used in SUNC to decrease the time differences effect was by employing a communication tool that enables sharing and exchanging information (JIRA). This tool had an extra advantage over face-to-face or synchronized video communication: the ability to record the conversations. This enabled distributed members to go back to the recording and take enough time to read and listen to the conversation. Participant D reported, "In addition to overlapping, we use a group chat tool that enables sending questions and feedback anytime."

Language Barrier: Speaking slowly, informing team members about the topics of discussion in the meeting, and addressing the language issue as early as possible in the project were good techniques used in SUNC to reduce the impact of language barriers. Moreover, all members had to use English during meetings. Participant E reported that

> We have a policy that whenever we have a session, all conversation should be in English during the session, which increases the effectiveness of the communication.

To make that possible, as a condition of joining SUNC, new team members must have excellent English. In addition, new team members had to attend training sessions, including English and technical terms used in SUNC. Participant E reported, "To overcome the language barrier, we have to run some training sessions in English for new members at the beginning when joining our teams."

Moreover, all members were required to participate and ask questions during the meetings, a practice that ensured that all members got the right message and practiced their English. The role of the product owner or scrum master on the Sydney team was important to ensure that all members participated in the meetings. As Participant D explained,

> You should be patient; take your time, and give enough time to every member; communicate clearly, make sure all members understand what the point of the message is, and when you are going through the meeting, give [everyone] the chance to talk. If you rush and do not take your time or you do not ask the question in the right way, you will have a misunderstanding of your message. You need to take away the barriers between you and them, so they talk.

Technologies and Tools: To mitigate the communication tools issue, SUNC continually

checked and updated their communication tools and ensured that they provided a high-speed data link. As Participant A explained, "We do a simulation of the tool twice a year to make sure it is [as] effective as [it is] supposed to be and to maintain the velocity of communication." Moreover, different communication tools (telephone, videoconference systems, instant messaging, email, and virtual whiteboards) were available so that if one failed, the other could be used. Participant E reported

> Technology is a big tip in establishing good communication, such as what we use here—"TV Link" and "chatting tool" that enable everyone to see all parts of the project [are] very helpful to communicate with distributed teams. Whatever we do (e.g., wall activities) here in Sydney [is] shared and accessible by all distributed teams, which keeps everyone aligned as much as possible.

Moreover, instant messaging (using JIRA or WhatsApp) among the three teams was used heavily during our observation of the Sydney team. Instant messaging was a good technique that compromised between real-time and asynchronous communication. It also gave the distributed team members the chance to understand the message, especially for the China and India team members who were non-native English speakers. Participant F reported,

> I personally prefer instant messaging when talking to distributed members. It keeps you connected with them all the time. It is better to understand each other.

Finally, some interviewees recommended adopting both formal and informal communication using the available tools. Although informal is preferable for agile, GDAD may need more formal communication, especially with different native spoken languages and multicultural teams. Participant D said that "Sometimes it is better to follow the formal channel with distributed teams to get the task done."

Personal Skills: Agile principles, such as scrum meetings and sprint planning, were recommended as tips to decrease the personal skills gap between distributed teams. For example, sprint planning meetings and reviews had a positive impact on the feature–requirements dependence because the use of these practices provided a systematic way to share information on the features and requirements among all distributed teams such as the testing teams, interface projects, and customers. Participant A stated

> I meet with the delivery leader, iteration manager, and scrum masters every two days. We go to the wall of cards (scrum of scrum) in addition to other meetings with the other team members to keep the track and scope alignment.

SUNC employed domain experts located in China or India, as reported by Participant G, "to decrease the differences in personal skills among our developers." This technique helped in solving some of the communication and coordination problems that occurred when remote teams were comprised of less experienced personnel than on the Sydney team. This practice enabled more scalability and increased trust between executive management and the distributed teams as the distributed teams came to better understand the business needs.

Documentation: In Sydney, members did not like to write too much documentation because they had the chance for face-to-face communication all the time. However, more documentation was needed to be shared with distributed teams. So as a procedure in SUNC, all members had to document the description of their code and point to any problem (technical or functional) related to the code. Also, they had to document what they had done and what they would do, and saved that on JIRA. Participant C reported,

> People, here [Sydney], do not like to read much documentation. So, we use less documentation within the onsite location. However, more documentation is still required with distributed teams.

This technique ensured that all members were up-to-date with what was done and what we needed, especially if some members no longer worked in Sydney or left the team. As Participant D emphasized, "Otherwise, how would we follow up?"

GDAD Communication Patterns Our in-depth analysis of the interviews and observations identified some communication patterns at SUNC, including practices and strategies. In the agile practice patterns, we noticed that all Sydney team members worked in an open office space. This gave them the opportunity for more informal communication, which increased the efficiency and effectiveness of the communication. Moreover, the open office environment provided a more efficient discussion and solution forum, on a daily basis, during the project. Participant B reported

> All the team members sit in this big office. We have an open channel with the project leads all the way. It has been a great success. It is much faster … to solve any issue here than work with distributed teams.

Furthermore, the product owner role was very important in facilitating communication among team members. They played the role of the real customer and contributed customer input to all other members. They were knowledgeable about the application domain and possessed the ability and skills to interact equally well with the technical team and business team members. They regularly communicated with the customer through videoconference systems, phones, emails, and face-to-face meetings. As Participant H described it, in team meetings (e.g., sprint planning), the product owner "was the key person who regularly [communicated] the new requirements. This speeded up communicating customer requirements among all teams."

In the strategic patterns, the Sydney site served as the hub for the distributed teams. The main site's members had significant experience working in SUNC and had completed several projects that involved other distributed teams (other than the India and China teams involved in the PBS project). This history gave them a great advantage in communicating with different cultures and languages. Communication between the China team and the India team was directed, in many cases, by the Sydney team. Conflicts between those two distant teams were often solved by the Sydney team, especially on requirement-related issues.

Furthermore, there was a common practice in SUNC that one member at the Sydney location spoke Urdu and another spoke Mandarin. This technique helped to facilitate communication between the main team in Sydney and the distributed teams in India and China. These members also regularly visited the corresponding distributed team during the project life cycle. This practice ensured that distributed teams remained on track.

Moreover, instant messaging (via JIRA and WhatsApp) was used heavily between the Sydney team and the distributed teams during our observation at the main site. This practice not only helped to enhance communication among team members but also culminated in documents that could be used in the future.

Discussion

The aim of this article is to answer the following two research questions.

> **RQ1.** What are the challenges of communication between GDAD teams?
> **RQ2.** How can the impact of GDAD communication challenges be mitigated?

Our findings have significant implications for researchers and practitioners of GDAD. We add insights to those presented in the existing literature about the most critical GDAD communication challenges and ways to prioritize those challenges, as well as revealing a new challenge (insufficient documentation). We reached these findings by conducting semistructured interviews and in-field observation of a large successful GDAD organization. Like any empirical study, this one has both implications and limitations, and we discuss them in the following sections.

Implications for Research and Practice ASD promotes informal communication using minimum documentation (e.g., using a backlog, card walls, and plans). These techniques have been shown to be highly successful in colocated teams. However, the situation is slightly different in GDAD. Our findings showed that one of the challenges facing GDAD communication was the insufficient documentation created by distributed teams. In large GDAD projects, due to many requirements, minimum documentation is very important [10] because without this documentation, team members may forget what they have done or become confused. Moreover, in some cases, the turnover in distributed members will be high and quick and those factors make it difficult for those joining the project to be effectively involved in a conversation with other distributed members or even with their own team members. This finding is consistent with the findings of Alzoubi and Gill [10], but this challenge has never been reported in previous studies by other researchers. In addition, this finding provides empirical evidence that documentation (to a reasonable level) should be provided in GDAD, thus helping practitioners to consider the necessary level of documentation to enhance communication and deliver a project successfully.

Although most previous empirical studies have shown that the major challenges of GDAD communication are time difference, difference of location, and language barriers, we show that the biggest challenge is cultural differences [49], as

reported by most of our interviewees. This represents another contribution of our study. Most previous research has not reported differences in the cultures of distributed members as a GDAD communication challenge. On the one hand, this finding demonstrates the need for more research to investigate this challenge in other GDAD firms and empirically sheds more light on techniques used to mitigate this challenge. On the other hand, GDAD practitioners need to pay more attention to this challenge, especially if different cultural teams or team members are involved.

Time differences, location differences, and language barriers were reported by several of our interviewees as GDAD communication challenges. These challenges are consistent with most of the previous studies. They have been reported for more than 10 years (see, e.g., Layman et al. [14]), thus showing that even after many years, these challenges still represent a major issue for GDAD. The recommendations or practices that we share in this article are similar to those proposed by previous studies, indicating that not many corrective practices have been adopted or perhaps that the practitioners and researchers have not been able to implement solutions. Consequently, further research and more innovative solutions are still needed. We provided one potential idea [10], recommending the use of agile enterprise architecture in GDAD that decreases the need for heavy communication among distributed teams and provides the blueprint of the whole project. This blueprint provides common ground and language that can be used among all distributed members and facilitates communication among them.

As discussed in the Research Context section, SUNC has been a successful GDAD firm because it has delivered many projects using the GDAD structure. Accordingly, the communication patterns used at SUNC can be an example for other GDAD firms. These patterns focus on the role of the product owner, who can facilitate communication with other teams and within the colocated team. Moreover, adopting a policy such as having representatives at the main location who can speak the native languages of the distributed teams had a positive impact during our observation of the Sydney location. Finally, the technique of training each new member at the Sydney location during the first month after joining the GDAD project could help decrease the impact of cultural and language differences. Hence, all of these patterns would be helpful to other GDAD firms.

Limitations There are some limitations of this article, as with any other empirical study, that should be considered when interpreting its findings. First, because collecting data using interviews is subjective by nature, we tried to decrease this effect by following an interview protocol during each interview. The protocol was developed based on the objectives of our study. We also triangulated interview data through field observation with the observation of the main team in Sydney.

Second, due to confidentiality constraints, accessibility to project work documents was rare, and that fact may affect the identification of communication challenges. Consequently, we analyzed as much as possible the documents that we could access in Sydney under the supervision of project team members.

Third, we faced some difficulties taking verbatim notes while conducting interviews, especially with distributed teams; hence, the majority of the interviews were conducted face-to-face with the Sydney team members.

Finally, the sample size is another limitation of this study. The standard sample size, according to Yin [43], is meaningless in the case study method. He also mentioned that researchers should concentrate on gathering information on different aspects of the case [43]. Therefore, there is no absolute minimum number of participants in a qualitative research project. The number of participants for a single case study is typically between 5 and 25 participants [50] and between 12 and 60 [51]. Accordingly, we interviewed 12 participants, thus meeting the required number. Furthermore, the aim of our study is to define and illustrate the case in depth rather than attempting to examine a number of different cases, so our results cannot be generalized to any other cases. Here, we used triangulation when collecting and interpreting data, which involves using various analysis techniques, data gathering methods, and data sources [51].

Conclusion

The success of ASD in a colocated context has motivated software development firms to adopt GDAD. However, GDAD faces many challenges. Communication has been identified as the main challenge of GDAD, resulting in delivering many unsuccessful projects. This lack of success results from the fact that ASD is a communication-oriented

approach; hence, insufficient (ineffective or inefficient) GDAD communication leads to delivery of projects that are late, exceed the budget, or fail to meet customer requirements. Several studies have tried to uncover the reasons for and solutions to this problem; however, there is still more to discover in this context. Accordingly, we have tried to shed more light on this issue by answering two research questions that focus on GDAD communication challenges and practices used to mitigate the impact of those challenges by in-depth analysis of semistructured interviews as well as the notes collected during our field observation of a successful firm that has adopted GDAD.

This article revealed six critical GDAD communication challenges: cultural differences, time differences, language barriers, communication tools, differences in personal skills, and insufficient documentation. Our findings revealed that the most critical challenge of GDAD communication is cultural differences among distributed teams. This finding has not been reported in previous literature as the most critical challenge. Moreover, our findings reveal that insufficient documentation is another important challenge of GDAD communication, in addition to the others that were identified in the previous literature. This finding emphasizes the importance of sufficient documentation in GDAD communication. The existence of these differences among distributed teams (distance, culture, and personal skills) makes "enough" documentation critical for achieving efficient and effective communication.

The findings of this article may urge GDAD practitioners to pay more attention to cultural differences among their distributed members as well as to ensure that they provide enough documentation. Moreover, the practices and recommendations provided in this article may guide GDAD firms in mitigating the effect of GDAD communication challenges reported here. Our findings may also promote more research on GDAD. This future research may focus on other practices used to mitigate the effect of these challenges in other successful GDAD firms, such as studying the various communication tools used among different firms.

References

[1] M. Shameem, R. R. Kumar, M. Nadeem, and A. A. Khan, "Taxonomical classification of barriers for scaling agile methods in global software development environment using fuzzy analytic hierarchy process," *Appl. Soft Comput.,* vol. 90, 2020, Art. no. 106122.

[2] R. Anwar, M. Rehman, K. S. Wang, and M. A. Hashmani, "Systematic literature review of knowledge sharing barriers and facilitators in global software development organizations using concept maps," *IEEE Access,* vol. 7, pp. 24231–24247, 2019.

[3] T. Kamal, Q. Zhang, M. A. Akbar, M. Shafiq, A. Gumaei, and A. Alsanad, "Identification and prioritization of agile requirements change management success factors in the domain of global software development," *IEEE Access,* vol. 8, pp. 44714–44726, 2020.

[4] J. F. Defranco and P. A. Laplante, "Review and analysis of software development team communication research," *IEEE Trans. Prof. Commun.,* vol. 60, no. 2, pp. 165–182, Jun. 2017.

[5] M. A. Babar, T. Ihme, and M. Pikkarainen, "An industrial case of exploiting product line architectures in agile software development," in *Proc. 13th Int. Softw. Product Line Conf.*, 2009, pp. 171–179.

[6] A. Yagüe, J. Garbajosa, J. Díaz, and E. González, "An exploratory study in communication in agile global software development," *Comput. Stand. Interf.,* vol. 48, pp. 184–197, 2016.

[7] R. A. Khan, M. Y. Idris, S. U. Khan, M. Ilyas, S. Ali, A. U. Din *et al.*, "An evaluation framework for communication and coordination processes in offshore software development outsourcing relationship: Using fuzzy methods," *IEEE Access,* vol. 7, pp. 112879–112906, 2019.

[8] AgileManifesto, Manifesto for agile software development. 2001. [Online]. Available: http://www.agilemanifesto.org/

[9] M. Shameem, R. R. Kumar, C. Kumar, B. Chandra, and A. A. Khan, "Prioritizing challenges of agile process in distributed software development environment using analytic hierarchy process," *J. Softw.-Evol. Proc.,* vol. 30, no. 11, 2018, Art. no. e1979.

[10] Y. I. Alzoubi and A. Q. Gill, "An empirical investigation of geographically distributed agile development: The agile enterprise architecture is a communication enabler," *IEEE Access,* vol. 8, pp. 80269–80289, 2020.

[11] S. Al-Saqqa, S. Sawalha, and H. AbdelNabi, "Agile software development: Methodologies and trends," *Int. J. Interact. Mob. Technol.,* vol. 14, no. 11, 2020, pp. 246–270.

[12] Y. I. Alzoubi, A. Q. Gill, and A. Al-Ani, "Empirical studies of geographically distributed agile development communication challenges: A systematic review," *Inf. Manage.,* vol. 53, no. 1, pp. 22–37, 2016.

[13] G. Ammad, U. I. Janjua, T. M. Madni, M. F. Cheema, and A. R. Shahid, "An empirical study to investigate the impact of communication issues in GSD in Pakistan's IT industry," *IEEE Access,* vol. 7, pp. 171648–171672, 2019.

[14] L. Layman, L. Williams, D. Damian, and H. Bures, "Essential communication practices for extreme programming in a global software development team," *Inf. Softw. Technol.,* vol. 48, no. 9,pp. 781–794, 2006.
[15] S. D. Müller, P. Kræmmergaard, and L. Mathiassen, "Managing cultural variation in software process improvement: A comparison of methods for subculture assessment," *IEEE Trans. Eng. Manage.,* vol. 56, no. 4, pp. 584–599, Nov. 2009.
[16] W. Aslam and F. Ijaz, "A quantitative framework for task allocation in distributed agile software development," *IEEE Access,* vol. 6, pp. 15380–15390, 2018.
[17] O. Turetken, I. Stojanov, and J. J. Trienekens, "Assessing the adoption level of scaled agile development: A maturity model for scaled agile framework," *J. Softw.-Evol. Proc.,* vol. 29, no. 6, 2017, Art. no. e1796.
[18] T. Dingsøyr, N. B. Moe, and E. A. Seim, "Coordinating knowledge work in multiteam programs: Findings from a large-scale agile development program," *Project Manage. J.,* vol. 49, no. 6, pp. 64–77, 2018.
[19] A. Martini, L. Pareto, and J. Bosch, "Communication factors for speed and reuse in large-scale agile software development," in *Proc. 17th Int. Softw. Product Line Conf.*, 2013, pp. 42–51.
[20] S. Dorairaj, J. Noble, and P. Malik, "Effective communication in distributed agile software development teams," in *Agile Processes in Software Engineering and Extreme Programming*, A. Sillitti, O. Hazzan, E. Bache, and X. Albaladejo. Berlin, Germany: Springer-Verlag, 2011, pp. 102–116.
[21] Y. I. Alzoubi, A. Q. Gill, and A. Al-Ani, "Distributed agile development communication: An agile architecture driven framework," *J. Softw.,* vol. 10, no. 6, pp. 681–694, 2015.
[22] P. J. Agerfalk, B. Fitzgerald, H. Holmstrom, B. Lings, B. Lundell, and E. O. Conchúir, "A framework for considering opportunities and threats in distributed software development," in *Proc. Int. Workshop Distrib. Softw. Develop.*, 2005, pp. 47–61.
[23] S. Carradini, "A comparison of research topics associated with technical communication, business communication, and professional communication, 1963–2017," *IEEE Trans. Prof. Commun.,* vol. 63, no. 2, pp. 118–138, Jun. 2020.
[24] S. Sarker, D. Nicholson, and K. D. Joshi, "Knowledge transfer in virtual systems development teams: An exploratory study of four key enablers," *IEEE Trans. Prof. Commun.,* vol. 48, no. 2, pp. 201–218, Jun. 2005.
[25] E. C. Brewer and T. L. Holmes, "Better communication = better teams: A communication exercise to improve team performance," *IEEE Trans. Prof. Commun.,* vol. 59, no. 3, pp. 288–298, Sep. 2016.
[26] P. E. Brewer, "Miscommunication in international virtual workplaces: A report on a multicase study," *IEEE Trans. Prof. Commun.,* vol. 53, no. 4, pp. 329–345, Dec. 2010.
[27] K. Crowston, J. Howison, C. Masango, and U. Y. Eseryel, "The role of face-to-face meetings in technology-supported self-organizing distributed teams," *IEEE Trans. Prof. Commun.,* vol. 50, no. 3, pp. 185–203, Sep. 2007.
[28] M. Hummel, C. Rosenkranz, and R. Holten, "The role of social agile practices for direct and indirect communication in information systems development teams," *Commun. Assoc. Inf. Syst.,* vol. 36, no. 1, pp. 273–300, 2015.
[29] E. Friess, "Scrum language use in a software engineering firm: An exploratory study," *IEEE Trans. Prof. Commun.,* vol. 62, no. 2, pp. 130–147, Jun. 2019.
[30] M. Pikkarainen, J. Haikara, O. Salo, P. Abrahamsson, and J. Still, "The impact of agile practices on communication in software development," *Empir. Softw. Eng.,* vol. 13, no. 3, pp. 303–337, 2008.
[31] P. Bjørn and O. Ngwenyama, "Technology alignment: A new area in virtual team research," *IEEE Trans. Prof. Commun.,* vol. 53, no. 4, pp. 382–400, Dec. 2010.
[32] Y. I. Alzoubi and A. Q. Gill, "Agile global software development communication challenges: A systematic review," in *Proc. 18th Pacific Asia Conf. Inf. Syst.*, 2014, p. 20.
[33] R. Paul, J. R. Drake, and H. Liang, "Global virtual team performance: The effect of coordination effectiveness, trust, and team cohesion," *IEEE Trans. Prof. Commun.,* vol. 59, no. 3, pp. 186–202, Sep. 2016.
[34] D. Anawati and A. Craig, "Behavioral adaptation within cross-cultural virtual teams," *IEEE Trans. Prof. Commun.,* vol. 49, no. 1, pp. 44–56, Mar. 2006.
[35] K. Siau, J. Erickson, and F. F.-H. Nah, "Effects of national culture on types of knowledge sharing in virtual communities," *IEEE Trans. Prof. Commun.,* vol. 53, no. 3, pp. 278–292, Sep. 2010.
[36] A. Alaiad, Y. Alnsour, and M. Alsharo, "Virtual teams: Thematic taxonomy, constructs model, and future research directions," *IEEE Trans. Prof. Commun.,* vol. 62, no. 3, pp. 211–238, Sep. 2019.
[37] Z. Guo, J. D'ambra, T. Turner, and H. Zhang, "Improving the effectiveness of virtual teams: A comparison of video-conferencing and face-to-face communication in China," *IEEE Trans. Prof. Commun.,* vol. 52, no. 1, pp. 1–16, Mar. 2009.
[38] D. Batra, D. VanderMeer, and K. Dutta, "Extending agile principles to larger, dynamic software projects: A theoretical assessment," *J. Database Manage.,* vol. 22, no. 4, pp. 73–92, 2011.
[39] Y. I. Alzoubi, A. Q. Gill, and B. Moulton, "A measurement model to analyze the effect of agile enterprise architecture on geographically distributed agile development," *J. Softw. Eng. Res. Develop.,* vol. 6, no. 1, 2018, Art. no. 4.
[40] R. M. Fuller, C. M. Vician, and S. A. Brown, "Longitudinal effects of computer-mediated communication anxiety on interaction in virtual teams," *IEEE Trans. Prof. Commun.,* vol. 59, no. 3, pp. 166–185, Sep. 2016.
[41] J. Suchan and G. Hayzak, "The communication characteristics of virtual teams: A case study," *IEEE Trans. Prof. Commun.,* vol. 44, no. 3, pp. 174–186, Sep. 2001.
[42] E. Bradner, G. Mark, and T. D. Hertel, "Team size and technology fit: Participation, awareness, and rapport in distributed teams," *IEEE Trans. Prof. Commun.,* vol. 48, no. 1, pp. 68–77, Mar. 2005.
[43] R. K. Yin, *Case Study Research: Design and Methods*. Thousands Oaks, CA, USA: Sage, 2009.

[44] Atlassian, What is Jira used for? 2020. [Online]. Available: https://www.atlassian.com/software/jira/guides/use-cases/what-is-jira-used-for
[45] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*. Beverly Hills, CA, USA: Sage, 1994.
[46] P. Mayring, "Qualitative inhaltsanalyse," in *Handbuch Qualitative Forschung in Der Psychologie*. New York, NY, USA: Springer, 2010, pp. 601–613.
[47] T. O'Donoghue and K. Punch, *Qualitative Educational Research in Action: Doing and Reflecting*. London, UK: RoutledgeFalmer, 2013.
[48] A. Fruhling and G.-J. De Vreede, "Field experiences with eXtreme programming: Developing an emergency response system," *J. Manage. Inf. Syst.,* vol. 22, no. 4, pp. 39–68, 2006.
[49] J. Iivari and N. Iivari, "The relationship between organizational culture and the deployment of agile methods," *Inform. Softw. Technol.,* vol. 53, no. 5, pp. 509–520, 2011.
[50] J. W. Creswell, *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. Newbury Park, CA, USA: Sage, 2012.
[51] S. E. Baker and R. Edwards, "How many qualitative interviews is enough," Nat. Centre Res. Meth., Southampton, UK, 2012. [Online]. Available: https://eprints.ncrm.ac.uk/id/eprint/2273/

Yehia Alzoubi received the M.Sc. degree in Information Technology from Central Queensland University Sydney, Rockhampton, QLD, Australia, and the Ph.D. degree in Information Management from the University of Technology Sydney, Sydney, NSW, Australia. He is an Assistant Professor with the Department of Management Information Systems, American University of the Middle East, Egaila, Kuwait. He has published in a number of international research journals and conferences, including *Information and Management*, the *Journal of Strategic Security*, *IEEE Access*, the Pacific Asia Conference on Information Systems, and the International Conference on Information Systems Development. His research interests include agile development, information security and privacy, and big data processing.

Asif Gill received the M.Sc. degree from the University of Newcastle Australia, Newcastle, Australia and the Ph.D. degree from the University of Technology Sydney, Sydney, Australia, both in Computing Science. He is an Associate Professor with the School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia. He has extensive professional IT experience in various sectors, including banking, consulting, software, and telecommunication. His work has appeared in major academic and industry software conferences and in information systems journals, including the *Australasian Conference on Information Systems*, *AMCIS*, the International Conference on Information Systems, the Pacific Asia Conference on Information Systems, the *Communications of the Association of Information Systems*, *Information Systems*, *Information and Management*, and *Information Systems Frontiers*. He is a coordinating editor, section editor, and editorial board member of several academic journals.

Research Article

"Better Communication Leads to a Higher Output?" An Analysis of Pair Communication on Pair Programming Productivity

—Stephen Choi

***Abstract*—Background:** *This study focuses on how group communication affects group productivity. The specific scope of the study is pair programming. The study aims to discover whether intra-pair communication in pair programming has a significant impact on the pair programming process and output.* **Literature review:** *Many of the pair programming communication studies are descriptive and qualitative studies whose foci lay more on communication contents and alternative message deliveries. As a result, more research that focuses on analyzing the effectiveness of a person's communication skill level while performing a demanding task is needed.* **Research question:** *Does the communication competency level significantly impact pair programming output?* **Methodology:** *A pool of novice university programming students was deployed for the experiment. The Conversational Skills Rating Scale (CSRS) was used to categorize them into three cohorts—"high-high," "high-low," and "low-low." The confounding variables were controlled.* **Results:** *No significant difference was found among the three cohorts in terms of their pair programming code output. Additionally, the post-experiment questionnaire responses revealed no significant difference in compatibility and confidence levels, but did show a significant difference in communication level.* **Conclusion:** *With all things being equal, a programmer's high communication skill level doesn't play a significant role in the programming output in a pair programming setting.*

Index Terms*—Communication, communication competency, pair programming.*

Among the list of attributes most employers, managers, and customers value in a business professional, the one that is probably most coveted is the ability to communicate effectively. Increasingly, high-technology organizations seek employees who are

> excellent communicators … who can sell the concepts of IT in creative ways with storytelling skills, translating the nuts and bolts to the decision-makers in terms they can understand. [1, p. 1]

For high profile social and mobile computing companies, securing this ability in their employees is seen as vital for their business sustainability as such companies must keep abreast of consumers' rapidly changing preferences and lifestyles. Being able to communicate with educated consumers and knowing what deliverables they want is a big part of the success for many high-tech companies. The significance of communication also applies to communication skills among employees. It is just as important for an employee to communicate effectively with a co-worker as it is with a customer. In short, communication is integral to all aspects of business and IT operations, especially because today's highly competitive market leaves little margin for error.

The agile software development (ASD) paradigm has been around for more than a decade and is now well established in the software engineering community [2]. With its famously unorthodox approaches to software and system development processes, it has attracted many headlines and brought radical changes [3] to mainstream software engineering practices as companies and employers seek higher efficiency [4]. The driving force of ASD lies with the following associated terms: Shared experience, human-centric, and communication-centric [2], [5]. ASD firmly believes in human and practical values (http://agilemanifesto.org). Communication is a central and vital process that dynamically integrates other components and parts

Manuscript received May 21, 2021; revised August 2, 2021; accepted August 3, 2021. Date of publication October 13, 2021; date of current version November 24, 2021. (*Corresponding author: Stephen Choi.*)
The author is with the Department of Information Systems and Decision Sciences, Craig School of Business,California State University Fresno, Fresno, CA 93740 USA (email: choi@csufresno.edu).
This paper has supplementary downloadable material at http://ieeexplore.ieee.org provided by the author. The file consists of two appendixes (66 kB in size).

IEEE 10.1109/TPC.2021.3110399

Practitioner Takeaway

- The study aims to discover whether intra-pair communication in pair programming has a significant impact on the pair programming process and output.
- Data analysis revealed that the high communication competency skill level has no significant influence on the pair confidence and compatibility.
- Under a school setting and all things being equal, a programmer's high communication skill level doesn't play a significant role in the programming output in a pair programming setting.

of this system [2], [6]–[8]. In the ASD environment, communication is the primary conduit for knowledge sharing and transfer. Consequently, the question of whether that conduit delivers a message in a manner anticipated by the receiver or whether it fails to do so must be addressed appropriately.

In terms of their research inquiries and studies, the research community has kept pace with the fast growth in the ASD sector. One particularly important ASD programming technique is what is called pair programming [2], [6]. To explain briefly, pair programming involves two programmers who code by sharing one keyboard and one monitor [9]. The idea behind two programmers sharing a single keyboard has been described using the phrase "two-heads are better than one" [10]. In particular, information sharing and pair effectiveness are central to this method's success. If the pair communicates effectively and often, then the communication bandwidth and frequency within the project will increase, thus causing greater information flow and team effectiveness [4]. However, this aspect of pair programming has led to challenges related to both cognitive and practical collaborations [11], [12].

Pair programing has been seen by some as possessing several shortcomings. Notably, these include the following.

1. Participants are not able to get into the "programming flow."
2. The "two for one" structure undermines productivity.
3. The pair is incompatible.
4. Many programmers are introverts [13].

Nonetheless, pair programing has established its niche within the software development community and is mostly seen as a perfectly viable choice. Most agree that it is most beneficial when pair programming is selectively or strategically implemented. For example, where there is minimal pressure, pair programming can flourish with the right pace and an ample amount of communication and interaction. This is typically seen, for instance, in the academic setting for novice programmers.

Given all these circumstances, communication appears to be a major dimension of pair programming. With that fact in mind, I searched for articles that specifically focused on pair programming communication but found only a few dozen studies that focused on the role of communication, content, behaviors, and disengagement [14]–[18]. Although these are substantial contributions, the fact remains that much more needs to be known about pair programming communication.

One important question is the role of the programmer's communication competency and its influence on the programming outcome. Communication competency can also be interpreted as efficient or effective communication. It is commonly assumed that more competent communicators are more likely to achieve their goals through communication. In industry, it is an axiom among organizations that they hold workshops or seminars to support effective internal communication to maintain or increase the productivity of their employees.

Many empirical studies support the view that higher levels of communication competency may have a positive effect on productivity and quality [19]–[21]. Sadia et al. [19] assessed how an employee's communication ability affects the overall internal communication within the organization and the employee's work capability and productivity. They report that effective communication culminates in a closer working relationship, a higher level of trust, and higher productivity. Clampitt and Downs [20] evaluated how the employee perceives the impact of communication satisfaction on productivity and

how the organizational type moderates the link between communication and productivity. Their study confirmed that the employee's perception of different communication variables affected their productivity. Furthermore, impact level and motives varied according to the variables.

When discussing communication and its impact on employee productivity, a few crucial questions arise. What is communication competency? Can it be measured? If so, how? Moreover, for the pair programming scenario, I must also ask how intrapair communication will be affected if both pair programmers are highly competent communicators? What if only one member of a pair is competent in communication? Another question that has not been addressed by any of the articles that I found is how does the level and competency of pair programming communication impact the coding output? The common expectation is that if programmers are highly competent in communicating, that competence will, in turn, yield greater coding output and a more satisfying pair programming experience. Nevertheless, two core questions have not been examined thoroughly enough.

- What will be the outcome if two highly communication-competent programmers are paired for programming compared to less optimal pairings?
- How does communication competency affect code productivity and code design?

Clearly this is a research gap that needs to be filled. Therefore, I present the following research question.

RQ. Does the communication competency level significantly impact pair programming output?

Tied into these questions is the fact that pair programming is a form of small group or dyad problem-solving and decision-making. Unlike other tasks, programming is probably one of the more demanding varieties of work; it requires a high level of cognitive processing [22]. Moreover, other facets of pair programming are closely aligned with small group problem solving [23]. Considering these facts, it would be prudent to evaluate pair programming communication through the perspectives of small group problem solving and decision making [24]. Not only is this a sensible approach, but it may provide us with new insights into the impact of communication on the problem solving, social cognition, and mutual understanding processes within pair programming. This approach has not been found in any of the previous pair programming articles and studies.

This study seeks to validate the relevance and significance of communication competency and its impact on pair programming. In addition, this study's results and its detailed analysis will provide the groundwork for subsequent pair programming communication studies. For practitioners, this study's results may provide another parameter to consider as they develop pair programming plans and or seek to find greater efficiencies.

Literature Review

Communication Perspectives Wyer and Adaval [25] explain communication conveyance through four categories: Literal, pragmatic, evaluative, and symbolic. In the *literal*, both sender and receiver use semantic meaning to construct their communication. In the *pragmatic*, either sender or receiver interprets with the aim of conveying his or her intentional meanings. In the *evaluative*, a receiver is faced with uncertainty and may interpret the meaning differently than intended by the sender. Finally, in the *symbolic*, a sender uses symbols to convey a message to a receiver.

The implications of this way of categorizing communication conveyance may be seen when one programmer is trying to communicate with the other programmer in an attempt to describe a certain concept or programming approach. The two must be "on the same page" in both understanding and collaborating. In a pair programming session, all categories of communication are possible as pair programming coding procedures and conventions, cognitive interactions, past programming experiences, decision-making, real-time code review, and other communications are taking place and being discussed. Furthermore, in pair programming or any other form of collaboration, disagreements and compromises are sure to be present. It is also important to remember that conflicts and compromises are not only subject to communication but also that communication is a key source of disagreements (when communication breaks down), as well as the means to finding their resolution.

A review of communication theories presents three schools of thought. The first is the trait model [26], [27]. The trait model speaks of communication competency as a dispositional or an innate characteristic. Just as some will see a person

naturally possessing an outgoing or a shy personality, the trait model sees good communication competency as a natural attribute of some individuals—one that is unlikely to change or fluctuate substantially over time.

In contrast, the second model predicts that communication competency can be acquired through repeated training sessions [28], [29]. According to this school of thought, communication competency is just another skill to be acquired and improved. Just like database administrator certification, network engineer certification, or other job-related skills, a person will ultimately achieve a certain level of communication competency for a job through repeated sessions of professional training.

The third model emphasizes social nurturing [30]–[33]. It acknowledges that communication competency relates to natural ability but also accepts that communication competency can be positively cultivated through a series of repeated social encounters over a long period. The social nurturing model emphasizes that repeated exposure and time are required for a person to cultivate "second-nature" competency.

Because of the exploratory nature of this study and the fact that there are no previous major pair programming communication studies to build on, this study opted for the trait model as its communication model. This decision reflects current pair programming practices where many programming shops provide no formal communication enhancement training programs for pair programmers but also may speak to the potential importance of providing such training.

Pair Programming Studies on Communication

The most common focus of previous pair programming studies is the behavior of the programmers during a session. These studies ask the following questions.

- How do the programmers communicate and interact with each other?
- Do they exhibit and support the "navigator" and "driver" roles and functions?
- What is the best theoretical framework to understand and construe the communication and behavior between the two?

As much as pair programming has shown its benefits and promise, many IT managers want to know about its inner operations and protocols to maximize its benefits.

In their experiment, Cao and Xu [14] controlled the programmer's programming ability to three tiers (high, medium, and low) and paired them in various formats. They reported that the high-high pairs exhibited more frequent "deeper thinking activities" than other pairs, and that the high-low pairs exhibited the lowest interactivity among all pairs. The satisfaction level assessment also mirrored the cognitive assessment findings. These findings are antithetical to the general understanding of pair programming benefits. Similary, Zarb et al. [18] presented a series of pair programming communication studies that utilized 31 pair programming session video recordings to code and classify the communications and communication behaviors of pair programmers. With no theory to reference, the studies used grounded theory [34]–[36]. The authors carefully analyzed and transcribed each video recording and reported the following codes: Suggestion, thinking, not focusing, explaining, reviewing, coding, and muttering.

Another pair programming behavior study by Freudenberg et al. [15] reported that the "navigator" and "driver" functions and roles did not hold well. They observed that the relationship resembles a lateral relationship rather than a hierarchical one (i.e., "like a cognitive tag team," [15, p. 85]) where they collaborate laterally. In addition, the theoretical underpinnings of pair programming communication are also lacking in this study. As a result, Hummel [16] called for theories that can lay the groundwork for pair programming communication that contributes to agile system development processes. Accompanying the communication is the cognitive aptitude of a programmer.

Another major dimension of pair programming is pair conflict. Domino et al. [37] presented two types of conflict in pair programming: Task conflict and affective conflict. Affective conflict refers to interpersonal conflict between the programmers, "interpersonal or affective conflict arises from feelings or emotions that are incompatible" [37, p. 46]. Task or goal conflict, on the other hand, occurs when the preferred outcomes between two parties appear to be incompatible. Instead of inquiring into why such conflicts happen, Domino focuses on how to manage such conflicts. Since conflicts and compromises are bound to happen in a typical pair programming session, the focus is on moderating the effects of the conflict-handling style, which includes taking into account the intrapair communication. In my review of the literature, I did not find a study that specifically focused on the

programmer's communication competency and its impact on the programmer's productivity.

Small Group Studies on Communication
Because problem solving and decision making are at the core of pair programming, pair programming can also be evaluated as another form of small group collaboration [38], [39]. Poole [40], [41] reported that in a group decision-making process, the phases do not necessarily follow in a sequential manner. The communication process often encounters "breakpoints" where new or different topics or strands develop, resulting in multiple strands of discourse, making the communication more unstructured, unpredictable, and multilayered.

In a small group problem-solving metaanalysis, Hirokawa [42]–[44] presented a set of five task achievement functions.

1. The group must establish a set of operating procedures.
2. The group must understand and analyze the problem.
3. The group must generate alternative solutions to solving the problem.
4. The group must develop a specific set of criteria for evaluating the worth of a given alternative solution.
5. The group must evaluate each alternative solution before making a final decision or identifying the solution.

The study reported that group effectiveness has a negative relationship with establishing operating procedure attempts, and a positive relationship with analyzing the problem attempts. The negatives of "establishing operating procedure" occur if there are excessive procedure-related comments and discussions. Such communication may be counterproductive and may impede the group problem-solving process, such as when a group spends too much time orienting and reorienting to the problem. As a result, they probably do not have a clear grasp of the problem itself, and consequently are not making progress toward solving the problem. Communicating too much about the pairing or programming procedures may undermine achieving the goal. However, sharing and verbalizing an appropriate amount of communication for the purposes of group problem-solving will enhance group process [45]. Thus, the key insight of this study is that spending too much of their interaction resources on negotiating procedural provisions is likely to hinder the pair programmers' progress.

Keyton et al. [46] studied the communication behaviors of working adults in an organizational setting where they categorized the worker's communication behaviors into the following categories: Information sharing, relational maintenance, expressing negative emotion, and organizing communication behaviors. I mapped these categories onto pair programming attributes.

- *Information sharing*—The programmers must actively share information and views.
- *Relational maintenance*—They must code in unison.
- *Expressing negative emotion*—The "navigator" must provide constructive feedback during real-time code review.
- *Organizing communication behaviors*—Verbal and nonverbal communication must be maintained.

A related study by Berger and Dibbattista [47] reported that when there is a misunderstanding or a communication failure, people do not change their discourse. Instead, they resort to different speed rates and vocal intensity (i.e., one often speaks more slowly and loudly for communication clarity rather than taking a different approach or perspective).

Communication Competency and Its Measurement "Communication competency" is a broad term and difficult to define [48] primarily because the term is defined differently according to each situational context. Currently, there is no definition for pair programming communication. The general understanding is that communication competency is influenced by a number of extrinsic and intrinsic variables such as culture, gender, context, purpose, proximity, and nonverbal cues. Out of these variables, probably the most elusive variables to determine are the nonverbal cues [49]. Birdwhistell [50] called nonverbal cues the "integrational aspects" of the communication process; they are defined as

> all behavior that keeps the system in operation, regulates the interaction process, cross-references particular messages to comprehensibility in a particular context, and relates the particular context to the larger contexts of which the interaction is but special situation. [50, p. 44]

The first prominent nonverbal cue is proximity. An ideal proximity is sometimes called the conversational "comfort zone." This is a psychological term, which describes the comfortable distance between sender and receiver

in holding a conversation without feeling too close to each other. When two individuals are sitting next to each other, this comfort zone also applies in a physical sense. The physical proximity of one's shoulders to the other's may indicate the level of mutually shared intimacy or friendship. Extrinsic variables such as gender may influence the comfort zone. For example, a female–female pair is more likely to be closer and more intimate with each other than a mixed gender one [51].

Besides the comfort zone, the next conspicuous nonverbal cue is the face. A person's emotion is relatively easy to read on their face. Ekman [52] reported that the face reveals emotion while the body reveals the intensity of emotion. In particular, the eye gaze provides more information than any other facial expression [53]. The intensity of eye gaze significantly influences a conversation as it indicates (among other things) how keen a person's attention is [54], [55]. The movements of other bodily parts may also relay important nonverbal cues.

Even with all the available nonverbal cues, measuring a person's communication competency is an elusive task [56]. The foremost question is what are the major items that need to be covered in measuring a person's communication competency in a pair programming context?

Cockburn [57] shared a list of communication cues between programmers.

- *Physical proximity*—The distance between the programmers; it also indicates tiny muscle movements, such as moving closer, which may indicate aggressiveness, enthusiasm, interest, agreement, or the desire to speak; or moving away, which may indicate fear, disagreement, or a private moment for a thought. The relative distance indicates the various levels of emotions as well as distinctive emotional status.
- *Smell*—Smell can be a factor when two programmers sit next to one another. Choi [58], for example, uses a field survey with a pool of experienced professional pair programmers to report how personal hygiene management is an issue to many pair programmers. It certainly is an influential factor in pair programming communication.
- *Kinaesthetics*—The sensation of movement to help us think and remember. The speaker might use it to help construct a new explanation or to improve the building of a question.
- *Touch*—Touch is a big part of communication as it symbolizes friendliness, supportiveness, and understanding. Using it appropriately and in a timely manner may ease tension and stress, and may induce a more favorable communication environment.
- *Sound*— A programmer may use colorful adjectives, exaggerations, and metaphors to make a point. The programmer may also control how he or she uses pitch, volume, and pacing when speaking a sentence to differentiate and emphasize a point.
- *Visuals*—Quite often, people use facial, hand, or body gestures to accentuate their points or views, such as raising an eyebrow, pointing while speaking, and raising their shoulders.

Embracing Cockburn's findings in the above list [57], I searched for a suitable communication competency measure instrument for pair programming communication. The search led to identifying conventional communication competency measurement instruments in the communication discipline. Among them, the most popular are the conversational skills rating scale (CSRS) form [59], [60]; the communicator competence questionnaire (CCQ) form [61]; the relational competence scale (RCS) form [32], [60]; and the dyadic adjustment scale (DAS) [62].

The next step was to determine which instrument was the most suitable for this study. After a careful review, I decided that the CSRS form was best suited (see Appendix A in the supplementary material). The CSRS consists of four subsections of behavioral items: Attentiveness (i.e., attention to, interest in, and concern for conversational partner), composure (i.e., confidence, assertiveness, and relaxation), expressiveness (i.e., animation and variation in verbal and nonverbal forms of expression), and coordination (i.e., the nondisruptive negotiation of speaking turns, conversational initiation, and conversational closings). Additional items that the four subsections of the behavioral items may have omitted were identified and added to the comprehensive measure. The additional items mainly evaluate the impressions of the partner: Whether he or she is a poor or good conversationalist, socially unskilled or skilled, an incompetent or competent communicator, an inappropriate or appropriate communicator, and an ineffective or effective communicator.

The items of the four subsections of behavioral items were identified as follows: Speaking rate,

speaking fluency, vocal confidence, articulation, vocal variety, volume, posture, leaning toward partner, shaking or nervous twitches, unmotivated movements or fidgeting, use of eye contact, facial expressiveness, nodding of the head in response to the partner's statements, use of gestures to emphasize what is being said, smiling or laughing, use of humor or stories, asking questions, encouragement or agreement, speaking about one's partner or a partner's interests, speaking about oneself, expression of personal opinions, initiation of new topics, maintenance of topics and follow up comments, interruption of a partner's speaking turns, and the use of time speaking relative to one's partner. All in all, it is a comprehensive list that matches closely with the behaviors observed in pair programming communication.

Hypothesis Development

Using the trait model of communication, I believe that programmers, who are highly competent communicators as measured by the CSRS form will communicate effectively and consequently generate a positive contribution to the coding output. Pair programming is comparable to small group problem solving, decision making, and collaboration; it is often accompanied by unstructured, unpredictable, and nonsequential conversation. Thus, programmers with higher levels of communication competency should be in a better position to deal with the dynamics of conversation than the less communication-competent programmers.

From the ASD side, several ongoing studies have placed an emphasis on communication in the ASD environment. Hummel et al. [63] and Hummel and Rosenkranz [7], [9] theorized that communication exerts a significant level of influence on information systems development (ISD) success. Pair programming is identified as a socially agile practice that puts a heavy emphasis on communication. In the studies, communication is divided into two subcategories: Communication informality (the degree of spontaneous and unstructured face-to-face conversations) and communication frequency (the amount of communication). Under this framework, higher levels of both communication informality and frequency are expected to yield a higher level of process performance quality and ISD success.

To view the matter from an alternative angle, I looked for similar approaches and phenomena in other domains. For instance, in business marketing, communication between a salesperson and a customer is evaluated to determine whether the salesperson's communication competency level impacts the sale or not [64]. As the customer searches for a particular item or tries to explain what item he or she requires, the salesperson assists the customer and joins in the search to enable mutual understanding. Or if the customer asks for more information or seeks a recommendation, the salesperson responds, explaining which item or information fits the customer's requirements. The customer and salesperson ask and respond to one another repeatedly, back and forth, until they agree on a particular item. In other words, they act out social cognition to comply with each other. The item may vary from a simple consumer good to a complex information technology product. Without a supporting framework of germane communication, they would not have reached their goal.

The situation of programmers involved in pair programming is comparable to that of the salesperson and customer. They similarly switch roles between "pilot" and "navigator." The programmer, like the salesperson, attempts to sell their coding idea or concept to the partner, the customer. They communicate like a salesperson and customer pair. When the roles are switched, the same salesperson-and-customer communication takes place again, but this time with the programmers taking on different roles. They share understanding to arrive at the final code. Thus, it stands to reason that a programmer's high level of communication competency will invite a higher level of intrapair communication during the role switches that may significantly contribute to a higher level of code output.

H1. The highly competent communicator paired with another person with similar competency will yield a significantly higher output in terms of both code productivity and code design than other pairs in which one or both individuals have a lower competency level.

The Hummel and Rosenkranz study [8] mentioned that better communication impacts the overall quality level of relationships among the ISD members in the following ways.

1. Team members share positive feelings for each other.
2. There is a sense of loyalty and responsibility.
3. There is a common goal.

These attributes eventually lead to a higher user satisfaction level.

Werner et al. [51], [65] reported that pair programmers who communicate well and who are well supported also show a strong sense of bonding and compatibility with each other. The highly competent communication pair will likely convey and illustrate their points in a clear and detailed manner. This will then lead to a higher level of compatibility and confidence than those pairs assessed at a lower communication competency level. In close, intimate dyadic relationships, communication is perceived to positively influence the relationship, whereas a lack of communication skills is likely to cause the relationship distress [66], [67]. So-called "rich" communication would also diffuse any pair differences or conflicts and bring them under control [68]. The communication construct is added to corroborate the validity and efficacy of the CSRS questionnaire for this context.

H2. Highly competent communication pairs will yield a significantly higher experience level in terms of compatibility, confidence, and communication than other pairs in which one or both individuals have a lower competency level.

Methodology

Experiment Process The experiment consists of two parts. The first part consists of

1. Holding an information session
2. Completing the participant consent form, questionnaire, and communication competency measurement session
3. Evaluating the participants' profiles for pairing
4. Scheduling the programming session dates
5. Notifying the corresponding pairs of the dates

The second part was the two programming sessions, with each consisting of a 45-minute pair-programming session. More detailed information is provided in the sections below.

Participants' Communication Competency Measurement From a pool of university students taking introductory programming courses, 144 participants joined the experiment. They had signed the study's consent form and were given course credit for their participation. In the orientation session, the participants were briefed about the experiment. They were then asked to fill out a programming background questionnaire to assess their suitability for the pairing process and to respond to the CSRS.

In this study, I assessed a participant's natural communication competency by using the CSRS. The facilitator asked the participants to pair-up with a "stranger" (another participant with whom they have had no prior personal relationship). Arbitrarily, every two participants were paired and a debating topic was given to each pair. The topic was related to a current situation affecting the university and two other nearby universities: The state government calling on its three state universities to merge into one large comprehensive state university. One university is predominantly an engineering school, the other is a medical school that also has a dental program, and the third is a liberal arts school that is a satellite campus of another state university. All three universities are located adjacent to each other. They allow their students to cross-register for courses and pursue a major at each other's university. The state government argues that this proposal is not only a sound economical move but that it will also boost and synergize the academic value of all three universities and their students. The opponents to this move argue that this move will "wash away" the long, rich academic traditions and uniqueness of the distinct universities that have made each special. The question of what would be the new name of this consolidated university was also a big issue. This topic was based on an actual case that was taking place during the time of this experiment. The administrators and students at each university dissented on the topic for many different reasons.

The pairs discussed the topic by clearly informing the partner of their own view, presenting their argument, and then listening to their partner's view and argument, before finally mitigating the differences. After a live discussion, the participants were asked to evaluate their partners using the CSRS form. The forms were collected for scoring and analysis.

From the 144 participants, 142 CSRS scores were acquired; two participants' scores were lost due to an administrative error. The scores were presented in a percentage format, where 100% indicated the highest communication competency according to the CSRS scoring system. For my study, I clustered the participants into three large groups: high, middle, and low. The next question was how to divide the participants into these three groups. I arbitrarily used the 0–40% score to create the low group, which yielded 44 participants, 41–59% for the middle group, which yielded 60 participants, and 60–100% for the high group, which yielded 38

participants. To address this arbitrary division and strengthen the competency measure, I dropped the middle group from the experiment, with the expectation that this change would create a more vivid contrast.

Pairing for the Experiment The focus of this pairing process was on controlling any confounding variables. Using the results of our earlier programming background questionnaire (to which the participants had given consent), I acquired information about each participant's academic performance in their current course, any prior programming courses they had taken as well as programming-related work outside the university, and information about their gender and academic major.

Regarding the participants' current academic performance level, the aim was to pair individuals with similar levels. Judgments regarding an individual's performance level were reached after intensive consultation with the programming course instructors. Prior programming experience was another possible factor. However, virtually none of the participants had taken any prior programming courses at the college level nor had any programming-related work experience. Only two students had taken an introductory programming course in high school, and they could not recall or identify any basic computer programming concepts. Thus, all the participants were considered to have no prior programming experience.

For the gender variable, a mixed-sex combination was used to control for that variable. When possible, I also cross-paired participants between sections to minimize the familiarity factor. If two participants were in the same section, they were asked whether they had any level of familiarity or friendship. Only when both participants' responses were negative were they allowed to pair up. The earlier CSRS instrument session was also subject to the familiarity check. If they were paired during the CSRS session, then they were not paired in this process.

With all participants' data juxtaposed, I carefully paired the participants while exercising due diligence to control those confounding variables. In the end, the study yielded 7 high-high [HH] pairs, 23 high-low [HL] pairs, and 7 low-low [LL] pairs. I lost 8 participants—1 high and 7 low—who did not qualify during the pairing process.

Experimental Instruments In developing the experimental problems, the focus was on the programming competency of the participants (novice level) and the difficulty level of the problems. Four outside professional programmers assisted in developing the problem set. The introductory programming course instructors were also consulted. Through a series of meetings and feedbacks, they optimized the problem set according to the profile of the participants and the scope of this study. Consequently, the problem set induced a reasonable amount of communication and interaction.

A postexperiment questionnaire was needed to capture the participants' experiment experience. No such questionnaire had existed before this study. Therefore, a new questionnaire was crafted. For the compatibility construct, three items were identified; for the confidence construct, four items were identified; and for the communication construct, 7 items were identified (see Appendix B in the supplementary material). The factor analysis used a principal components analysis with varimax rotation. The parameters were eigenvalues greater than one, and maximum iteration for convergence was set to be 25. Any item with a value less than 0.500 was discarded, as was any construct with less than the minimum three items. The Cronbach's alpha measurements were 0.7084 for compatibility, 0.8314 for confidence, and 0.8794 for communication.

The experiment started by informing the participants about their visiting hours and dates. Two pair programming sessions were scheduled for each pair, with one problem set per session. An experiment facilitator hosted the sessions in a controlled computing laboratory. For each session, 45 minutes were allocated to solve and finish the problem. Even if the problem was not completely solved within this time, the pairs were asked to turn in their work. All the sessions were video recorded to capture the communication, seating proximity, facial expressions, body movements, and pilot and navigator role switching of the participants. The video recordings facilitated a better understanding of the pair programming sessions.

Judging All the pairs' coding outputs, including all error messages, were collected and given to a panel of two outside judges. Both judges hold advanced degrees in computer science and have extensive programming experience. They were briefed on the study's background and

TABLE I
KOLMOGOROV–SMIRNOV TEST RESULT 1

Dependent Variables	Statistic	Degrees of Freedom	Significance Probability
Code Productivity	0.145	123	0.000
Code Design	0.118	123	0.000

experimental processes. Code productivity refers to how much meaningful coding was completed, and code design refers to code efficiency and readability. With a small sample of coding output and a few practice coding sessions, the two judges reconciled their subjective assessment differences.

The grading scale was 0 to 10, with 0 being the lowest score and 10 being the highest score. If there was a difference of more than 1 between the two judges' scores, they were asked to reconcile for adjustment. A bivariate correlation was performed to assess the interjudge reliability—a value of 0.963 for code productivity and a value of 0.967 for code design—which indicated that the two judges' assessments were highly correlated.

CONTENT ANALYSIS

Data Normal Distribution Check To ensure the integrity of this study's data analysis and its results, I administered the data normal distribution check [69], which corroborates whether the dataset is normally distributed and clears the dataset for parametric tests [69]. The visual inspection tools of a dataset, such as histogram or Q-Q graph, tend to provide a more subjective evaluation and could discredit the result. Therefore, performing a test (once or more) on the data's normal distribution will bolster the dataset's integrity. The descriptive statistics results of the dependent variables code productivity and code design yielded Skewness values of −0.320 and −0.388, respectively. For normal data distribution, the Skewness value must be zero. Hence, these values indicate that the data does not pass the normal distribution check. Table I shows the Kolmogorov–Smirnov test result. That test shows the p values of 0.000 in both dependent variables, which indicate that both dependent variable datasets are not normally distributed.

Table II shows the dependent variable mean values among the three cohorts: High-high pairs [HH], high-low pairs, [HL], and low-low pairs [LL]. Surprisingly, the [LL] had the highest mean value, and the [HH] had the lowest. The [HH] also had the highest standard deviation values.

TABLE II
PROFILES OF THE THREE COHORTS

Dependent Variables	Pair Category	Pair *N*	Mean	*SD*
Code Productivity	[HH]	7	4.79	3.33
	[HL]	23	5.78	2.83
	[LL]	7	5.97	2.72
Code Design	[HH]	7	5.07	3.01
	[HL]	23	5.64	2.74
	[LL]	7	5.75	2.94

TABLE III
KRUSKAL–WALLIS TEST RESULT

	Code Productivity	Code Design
Chi-Square	1.330	0.369
Degrees of Freedom	2	2
Significance Probability	0.514	0.832

To determine whether the mean values in Table II are significant or not, the Kruskal–Wallis test was administered (see Table III). The Kruskal–Wallis test is a nonparametric test for a dataset that is not normally distributed, like the one-way ANOVA test with a normally distributed dataset. The p values of both code productivity and design are 0.514 and 0.832, respectively, indicating they are not significantly different from each other. In other words, the levels of code productivity and design are not significantly influenced by the communication competency levels of the pair programmers.

Thus, in reviewing the study's results for H1,

> **H1.** The highly competent communicator paired with another person with similar competency will yield a significantly higher output in terms of both code productivity and code design than other pairs in which one or both individuals have a lower competency level.

we find that H1 is not supported for either code productivity output or code design output.

POSTEXPERIMENT QUESTIONNAIRE ANALYSIS

Data Normal Distribution Check Similar to the data normal distribution check in the content analysis, the normal distribution check was also administered with the postexperiment questionnaire datasets. The descriptive statistics results for compatibility, confidence, and

TABLE IV
KOLMOGOROV–SMIRNOV TEST RESULT 2

Constructs	Statistic	Degrees of Freedom	Significance Probability
Compatibility	0.121	210	0.000
Confidence	0.059	210	0.076
Communication	0.112	210	0.000

TABLE V
CRONBACH'S RELIABILITY MEASUREMENT

Constructs	Reliability (Alpha)
Compatibility	0.7084
Confidence	0.8314
Communication	0.8794

communication yielded Skewness values of −0.879, −0.646, and −0.966, respectively. Additionally, the Kolmogorov–Smirnov test was administered. Table IV shows the *p* values of 0.000, 0.076, and 0.000 for compatibility, confidence, and communication. For a normal distribution, the *p* value must be greater than 0.05. Only the confidence construct showed the normal distribution. For a collective data analysis and consistency, all three constructs were treated with nonparametric test for the next analysis.

The Crobach's Alpha measurement for each set is above 0.700; see Table V. For the communication construct, the value is 0.8794, which confirms its reliability. All three constructs' values are above 0.70.

Data Analysis Equivalent to the ANOVA in its purpose, the nonparametric Kruskal–Wallis test was conducted (see Table VI). The *p* values for compatibility, confidence, and communication constructs were 0.643, 0.671, and 0.040, respectively. A *p* value of 0.050 or less indicates a significance difference. Only the communication construct exhibited a significant difference.

With the above Kruskal–Wallis test result, only the communication construct was analyzed further. A nonparametric test, the Mann–Whitney test, was used to determine any significant differences between the three cohorts. Table VII shows the result. Between [HH] and [HL], there is a significant difference with a *p* value of 0.027, and between [HH] and [LL], there is a significant difference with *a* p value of 0.008, but there is no significant difference between [HL] and [LL], with a *p* value of 0.115.

Thus, in reviewing the study's results for H2,

> **H2.** Highly competent communication pairs will yield a significantly higher experience level in terms of compatibility, confidence, and communication than other pairs in which one or both individuals have a lower competency level.

we find that H2 is supported only for the communication construct.

DISCUSSION

This study's results demonstrate that having two pair programmers with high levels of communication competency did not necessarily positively influence code productivity or code design. In addition, the results did not positively influence the levels of compatibility and confidence between the pair programmers. It supported only the validity of the communication competency measurement that was used in this study, which was based on the trait model.

Table II shows that the mean value of both code productivity and the design of the high-high pairs was the lowest. The standard deviation value for [HH] was the largest of the three. Therefore, H1 was not supported. The common belief that one who communicates well also contributes positively to group collaboration does not stand up in the pair programming context. The "rich" communication or "more" communication thesis does not directly link to a better outcome or greater output.

The outcome is in stark contrast to the general findings and consensus about effective communication in working relationships [19]–[21]. A few plausible explanations are as follows.

- Participants were not able to get into the "programming flow."
- Most programmers are introverts [13].

It is also important to remember that the cognitive intensity and challenge are quite high (in many cases, at the maximum level) for most programming tasks. Therefore, it is necessary that one get into the "programming flow" to successfully finish a programming task. This can be difficult in a pair programming context where two programmers must get in the "programming flow" through their communication. Moreover, for an introverted programmer, it must be difficult to

TABLE VI
KRUSKAL–WALLIS TEST RESULT ON THE THREE CONSTRUCTS

	Compatibility	Confidence	Communication
Chi-Square	0.884	0.798	6.436
Degrees of Freedom	2	2	2
Asymptotic Significance	0.643	0.671	0.040

TABLE VII
MANN-WHITNEY TEST ON THE THREE COHORTS

	[HH] vs. [HL]	[HH] vs. [LL]	[HL] vs. [LL]
Mann-Whitney U	763.0	179.000	824.5
Wilcoxon W	4166.0	479.000	1124.5
Z	-1.93	-2.420	-1.204
Significance	0.027	0.008	0.115

communicate effectively to conform to the pair programming protocols [70]. Traditionally, programming is viewed as an individualized activity with limited interaction with others, but this case is the opposite as a result of using pair programming protocols.

Another explanation is the inherent nature of pair programing procedures and what setting them up entails. From the small group problem-solving perspective, Hirokawa [42]–[44] mentions that group effectiveness is negative when too much effort is spent establishing operating procedures. On the other hand, positive group effectiveness is achieved when efforts are channeled toward analyzing the problem. Translated into pair programming terms, if the highly communication competent pair programmers spend too much time and effort talking about pair programming protocols or other related peripheral issues, rather than discussing the actual problem and possible solutions, then the pair programming session would not be efficient.

On the other hand, if the highly competent communication pair programmers did spend their time and efforts analyzing the problem, then the mean score should have been significantly higher than it was [45]. Under the circumstances, this finding may mean that the programmers spent too much time and effort in seeking alternative problem-solving approaches. This behavior can be considered subservient to establishing operating procedures. To ascertain the type of actual communication and its content, the video recorded pair programming sessions would need to be examined. Specifically, each comment and utterance of all participants would need to be coded for further analysis. This task will be conducted in a subsequent study, which will shed more light on pair programming communication as well as add substantively to the discussion. For this coding process, I expect to reference and utilize the studies of Poole [40], [41]; Hirokawa [42]–[44]; and Keyton et al. [46].

From a communication perspective, this study proved that a high level of natural communication competency does not necessarily have a positive influence on pair programming output. The validity of this finding is supported by our results for Hypothesis 2. The level of the communication competency measurement before and after the experiment was validated. However, this finding casts doubt on the use and effectiveness of the trait model for the study of pair programming communication. Although the receiver acknowledged the richness of communication, that richness did not translate into an effective programming process. The above argument about whether the communication type was either establishing operating procedures or analyzing the problem is also tied to this issue. The communication type could have hampered the potency of the communication competency level. Alternatively, I may shift the focus to the communication perspective that states a person's communication competency level can be enhanced through repeated professional training sessions, as well as through a process of social nurturing.

In terms of the pair programming experience of the participants, the insignificance of the compatibility and confidence levels between the three cohorts may have been caused by the insignificance of the code output levels between the three cohorts. In other words, the participants may not have been satisfied because they were unable to reach their goals. Another plausible explanation relates to the communication type again: Establishing operating procedures or analyzing the problem. The

participants could have been more satisfied with one type over another. If the participants were more focused on completing the coding task, then they would have been significantly more satisfied with analyzing the problem communication type.

Yet another issue is the possibility of dyad conflicts. Generally, computer programming is known for its demanding work. As it becomes more demanding, the chance of dyad conflicts will increase—especially when the two heads are supposed to function as one, as in pair programming. There may also be other types of conflicts to consider: Those that are task-related—procedural and cognitive, and affective or personal. The term "procedural conflict" refers to a conflict type where the two programmers experience differences with respect to pair programming protocol [12]. Cognitive conflict refers to a conflict type, where two programmers differ in coding or problem-solving approaches [22], [23], [71]. Lastly, personal conflict refers to a situation in which two programmers experience personal differences, such as those relating to personality or gender [51], [72], [73]. These different types of conflict are to be controlled and managed. In face-to-face pair programming, communication is the main conduit between the programmers that can address those conflicts. From this fact, we may infer that there could be a closer relationship between conflict management and pair programming communication.

The communication construct was significantly different among the three cohorts. This fact clearly proves that the communication competency measurement is valid and effective. But more important, it illuminates an intriguing interaction effect among the compatibility, confidence, and communication constructs. The fact that the participants experienced significantly different communication levels but no significantly different compatibility and confidence levels suggests important meanings. Moreover, this fact repudiates another general perception: If a pair of individuals is engaged in a "rich" form of communication, that rich communication does not necessarily bring with it high levels of compatibility and confidence. The key point to remember is that this argument deals with the context of pair programming or dyad problem solving.

Limitations There are important limitations to mention. The first limitation is the small size of the dataset: Only seven high-high [HH] pairs, 23 high-low [HL] pairs, and seven low-low [LL] pairs were assessed. This limitation is largely the result of eliminating participants with communication competency at the middle level, and the pairing of the participants in controlling the confounding variables. Another limitation is that the participant pool was comprised of university students and not professional programmers. Experienced, well-versed programmers may position themselves better in managing the pair programming session than the university students, and this better management would likely influence the outcome. The results of this study must be interpreted with these circumstances in mind.

Contributions and Future Studies In terms of contributions, this study empirically tested how a programmer's natural communication competency affected pair programming output and experience. This was one of the basic questions about pair programming communication. The results of this study provide the next set of questions, which are described in the following paragraph in more detail. The results of this study also provide the facts that will help steer future studies regarding this topic, including potentially redefining the role and implications of intrapair communication in pair programming. In terms of practical contribution, this result will assist programming managers in re-evaluating their priorities when pairing programmers for pair programming sessions. Lastly, this study adds to the collective efforts of IS researchers who are studying the impact of communication on pair programming or other ASD methods.

For future studies, the most important step is to code the video recordings of pair programming sessions. This information should reveal significant insights into which communication type is inclined more toward establishing procedures or analyzing problem-solving options. Previous studies and code schemes have looked at communication content with a more objective measure and analysis [74], but I will evaluate this topic with greater attention to different contexts and situations. Furthermore, the coded information will be analyzed and evaluated in a more detailed manner by examining small group problem-solving and decision-making collaboration. This focus may reveal hidden idiosyncrasies and saliencies of dyad dynamics in pair programming. In addition to communication, I will also focus on conflict management because it is perceived to have a major effect on pair programming. During coding, I will identify and classify conflicts, according to the type, frequency, duration, and subsequent communication and

behavior. Eventually, the focus will be on how these conflicts impact the pair programming process and its coding output.

Moreover, nonverbal communication will be another major factor in this next study. I will record and classify the various forms of nonverbal communication, including seating proximity, facial expressions, eye gaze, hand gestures, shoulder positioning, speech duration, and other significant cues. In relation to verbal communication, I will identify the nonverbal communication that accompanies positive communication, such as compatibility and confidence, as well as nonverbal communication that results in negative communication, such as experiencing a conflict.

Another study that I have planned will use an instrument to measure the level of conflict. A better understanding of pair programming conflicts and their mechanisms would bring more information from the conflict management dimension. Taking a more microscopic view than that currently used, one instrument that may be well suited to future studies is the Conflict Resilience Quotient form, which measures resiliency to group conflicts [75].

Combining this information, I may be able to discover a great deal about the nature of pair programming communication. This may also provide a starting point to formulating a framework for pair programming communication.

Acknowledgment

The author would like to thank Dr. B. H. Spitzberg, Senate Distinguished Professor Emeritus, School of Communication, San Diego State University, for his assistance, comments, and advice on an early draft of this article.

References

[1] C. Hall, "Creative employees in high demand among new tech hires," *Forbes*, Aug. 2021. [Online]. Available: http://www.forbes.com/sites/emc/2013/07/24/creative-employees-in-high-demand-among-new-tech-hires/print/
[2] A. Cockburn, *Agile Software Development*. 2nd ed. Boston, MA, USA: Pearson Education, 2006.
[3] R. L. Glass, "Extreme programming: The good, the bad, and the bottom line," *IEEE Softw.*, vol. 18, no. 6, pp. 112–121. Nov./Dec. 2012.
[4] A. Cockburn and L. Williams, "The costs and benefits of pair programming," in *Extreme Programming Examined*, G. Succi and M. Marchesi, Eds. Boston, MA, USA: Addison-Wesley Longman Publishing, 2013, pp. 223–247.
[5] A. Cockburn and J. Highsmith, "Agile software development, the people factor," *Computer,* vol. 34, no. 11, pp. 131–133, 2001.
[6] J. A. Highsmith, *Agile Software Development Ecosystems*, Boston, MA, USA: Addison-Wesley, 2002.
[7] M. Hummel and C. Rosenkranz, "Is communication the key to success? Investigating the impact of agile practices on information systems development projects," in *International Research Workshop On IT Project Management*, 2012, pp. 138–146.
[8] M. Hummel and C. Rosenkranz, "Measuring the impact of communication in agile development: A research model and pilot test," in *Proc. 19th Americas Conf. Inf. Syst.*, 2013, pp. 1–10.
[9] L. Williams and R. R. Kessler, *Pair Programming Illuminated*, Boston, MA, USA: Addison-Wesley, 2002.
[10] V. Balijepally, R. Mahapatra, S. Nerur, and K. Price, "Are two heads better than one for software development? The productivity paradox of pair programming," *MIS Quart.*, vol. 33, no. 1, pp. 91–118, 2009.
[11] V. Balijepally, R. Mahapatra, and S. Nerur, "Cognitive dispositions of IS personnel and implications for pair programming," in *Proc. 9th Americas Conf. Inf. Syst.*, 2003, pp. 1532–1539.
[12] L. Williams and R. R. Kessler, "All I really need to know about pair programming I learned in kindergarten," *Commun. ACM*, vol. 43, no. 5, 2000, pp. 108–114.
[13] J. Thomas, "Love it or hate it, pair programming is here to stay. Here's how you can live with it," Coscreen, Menlo Park, CA, USA, Jan. 5, 2021. [Online]. Available: https://www.coscreen.co/blog/love-or-hate-pair-programming-here-to-stay/
[14] L. Cao and P. Xu, "Activity patterns of pair programming," in *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.*, 2005, pp. 1–10.
[15] S. Freudenberg, P. Romero, and B. du Boulay, "Talking the talk: Is intermediate-level conversation the key to the pair programming success story?" in *Proc. Agile Conf.*, 2007, pp. 84–91.
[16] M. Hummel, C. Rosenkranz, and R. Holten, "The role of communication in agile systems development." *Bus. Inf. Syst. Eng.*, vol. 5, no. 5, pp. 343–355, 2013.
[17] L. Plonka, H. Sharp, and J. van der Linden, "Disengagement in pair programming: Does it matter?" in *Proc. 34th Int. Conf. Softw. Eng.*, 2012, pp. 496–506.
[18] M. Zarb, "Understanding communication within pair programming," in *Proc. 3rd Annu. Conf. Syst., Program., Appl.: Softw. For Humanity*, 2012, pp. 53–55.
[19] A. Sadia, B. M. Salleh, Z. A. Kadir, and S. Sanif, "The relationship between organizational communication and employees productivity with new dimensions of effective communication flow," *J. Bus. Social Rev. Emerg. Economies*, vol. 2, no. 2, pp. 93–100, 2016.

[20] P. G. Clampitt and C. W. Downs, "Employee perceptions of the relationship between communication and productivity: A field study," *J. Bus. Commun.*, vol. 30, no. 1, pp. 5–28, 1993.
[21] I. Opitz and M. B. Hinner, "Good internal communication increases productivity," *Freiberger Arbeitspapiere*, vol. 7, 2003. [Online]. Available: https://www.econstor.eu/bitstream/10419/48382/1/369161947.pdf
[22] G. M. Weinberg, *The Psychology of Computer Programming*. New York, NY, USA: Dorset House, 1998.
[23] L. L. Constantine, *Constantine On Peopleware*. Upper Saddle River, NJ, USA: Yourdon Press, 1995.
[24] J. E. McGrath, *Groups: Interaction and Performance*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
[25] R. S. Wyer and R. Adaval, "Message reception skills in social communication," in *Handbook of Communication and Social Interaction Skills,* J.O. Greene, B.R. Burleson Eds. Mahwah, NJ, USA: Erlbaum, 2002, pp. 291–355.
[26] S. W. Littlejohn and K. A. Foss, *Theories of Human Communication*, 9th ed. Belmont, CA, USA: Cengage Learning, 2008.
[27] J. M. Wiemann, "Explication and test of a model of communicative competence," *Human Commun. Res.*, vol. 3, no. 3, pp. 195–213, 1977.
[28] B. R. Burleson, "Constructivism: A general theory of communication skill," in *Explaining Communication: Contemporary theories and Exemplars*, B.B. Whaley, W. Samter, Eds. Mahwah, NJ, USA: Erlbaum, 2007, pp. 105–128.
[29] W. C. Sharbrough, S. A. Simmons, and D. A. Cantrill, "Motivating language in industry: Its impact on job satisfaction and perceived supervisor effectiveness," *J. Bus. Commun.*, vol. 43, no. 4, pp. 322–343, 2006.
[30] D. P. Cushman and R. T. Craig, "Communication systems: Interpersonal implications," in *Explorations in Interpersonal Communication*, G.R. Miller, Ed. Beverly Hills, CA, USA: Sage, 1976, pp. 37–58.
[31] R. M. McFall, "A review and reformulation of the concept of social skills," *Behav. Assessment*, vol. 1, no. 4, pp. 1–33, 1982.
[32] B. H. Spitzberg and W. R. Cupach, *Interpersonal Communication Competence*. Beverly Hills, CA, USA: Sage, 1984.
[33] B. H. Spitzberg and W. R. Cupach, *Handbook Or Interpersonal Competence Research*, New York, NY, USA: Springer-Verlag, 1989.
[34] M. Birks and J. Mills, *Grounded Theory: A practical Guide*. Thousand Oaks, CA, USA: Sage, 2011.
[35] S. Salinger, L. Plonka, and L. Prechelt, "A coding scheme development methodology using grounded theory for qualitative analysis of pair programming," *Human Technol.*, vol. 4, no. 1, pp. 9–25, 2008.
[36] A. Strauss and J. Corbin, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London. UK: Sage, 1990.
[37] M. A. Domino, R. W. Collins, A. R. Hevner, and C. F. Cohen, "Conflict in collaborative software development," in *Proc. SIGMIS Conf. Comput. Personnel Res.*, 2003, pp. 44–51.
[38] A. J. Salazar, "Ambiguity and communication effects on small group decision-making performance," *Human Commun. Res.*, vol. 23, no. 2, pp. 155–192, 1996.
[39] H. Witteman, "Interpersonal problem solving: Problem conceptualization and communication use," *Commun. Monographs,* vol. 55, no. 4, pp. 336–359, 2009.
[40] M. S. Poole, "Decision development in small groups II: A study of multiple sequences in decision making," *Commun. Monographs*, vol. 50, no. 3, pp. 206–232, 1983.
[41] M. S. Poole, "Decision development in small groups, III: A multiple sequence model of group decision development," *Commun. Monographs*, vol. 50, no. 4, pp. 321–341, 1983.
[42] R. Y. Hirokawa, "Group communication and problem-solving effectiveness II: An exploratory investigation of procedural functions," *Western J. Commun.*, vol. 47, no. 1, pp. 59–74, 1983.
[43] R. Y. Hirokawa, "Group communication and problem-solving effectiveness: An investigation of group phases," *Human Commun. Res.*, vol. 9, no. 4, pp. 291–305, 1983.
[44] R. Y. Hirokawa, "Discussion procedures and decision-making performance," *Human Commun. Res.*, vol. 12, no. 2, pp. 203–224, 1985.
[45] L. Wilder and D. J. Harvey, "Overt and covert verbalization in problem solving," *Commun. Monographs*, vol. 38, no. 3, pp. 171–176, 1971.
[46] J. Keyton *et al.*, "Investigating verbal workplace communication behaviors," *J. Bus. Commun.*, vol. 50, no. 2, pp. 152–169, 2013.
[47] C. R. Berger and P. Dibattista, "Communication failure and plan adaptation: If at first you don't succeed, say it louder and slower," *Commun. Monographs,* vol. 60, no. 3, pp. 220–238, 1993.
[48] S. R. Wilson and C. M. Sabee, "Explicating communicative competence as a theoretical term," in *Handbook of Communication and Social Interaction Skills,* J.O. Greene, B.R. Burleson, Eds. Mahwah, NJ, USA: Erlbaum, 2003, pp. 3–50.
[49] P. Ekman, "Communication through nonverbal behavior: A source of information about an interpersonal relationship," in *Affect, Cognition and Personality,* S.S. Tomkins, C.E. Izard, Eds. New York, NY, USA: Springer, 1965, pp. 390–3442.
[50] R. L. Birdwhistell, *Kinesics and Context*. Philadelphia, PA, USA: Univ. of Philadelphia Press, 1970.
[51] L. L. Werner, B. Hanks, and C. McDowell, "Pair programming and gender," in *Encyclopedia of Gender and Information Technology*, E.M. Trauth, Ed. Hershey, PA, USA: Idea Group, 2006, pp. 957–963.
[52] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry,* vol. 32, no. 1, pp. 88–105, 1969.
[53] A. Kendon, "Some functions of gaze direction in social interaction," *Acta Psychol.*, vol. 26, pp. 22–63, 1967.
[54] M. Argyle and J. Dean, "Eye contact, distance and affiliation," *Sociometry*, vol. 28, no. 3, pp. 289–304, 1965.
[55] G. N. Goldberg, C. A. Kiesler, and B. E. Collins, "Visual behavior and face-to-face distance interaction," *Sociometry*, no. 32, no. 1, pp. 43–53, 1969.

[56] F. M. Jablin and P. M. Sias, "Communication competence," in *New Handbook of Organizational Communication: Advances in Theory, Research, and Methods*, F.M. Jablin and L.L. Putnam, Eds. Thousand Oaks, CA, USA: Sage, 2011, pp. 819–864.
[57] A. Cockburn, *Agile Software Development*. Boston, MA, USA: Addison-Wesley, 2001.
[58] K. Choi, "Team programming influencing factors: A field survey," *J. Int. Technol. Manage.*, vol. XVIII, no. 3/4, pp. 1–13, 2007.
[59] B. H. Spitzberg and H. T. Hurt, "The measurement of interpersonal skills in instructional contexts," *Commun. Educ.*, vol. 36, no. 1, 1987, pp. 28–45.
[60] B. H. Spitzberg, "Communication competence: Measures of perceived effectiveness," in *Handbook For the Study of Human Communication*, C.H. Tardy, Ed. Norwood, NJ, USA: Ablex, 1988, pp. 67–105.
[61] P. R. Monge, S. G. Bachman, J. P. Dillard, and E. M. Eisenberg, "Communicator competence in the workplace: Model testing and scale development," in *Communication Yearbook*, M. Burgoon, Ed. New Brunswick, NJ, USA: Transaction, 1982, pp. 505–528.
[62] G. B. Spanier, "Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads," *J. Marriage Fam.*, vol. 38, no. 1, pp. 15–28, 1976.
[63] M. Hummel, C. Rosenkranz, and R. Holten, "Explaining the changing communication paradigm of agile," in *Proc. 21st Eur. Conf. Inf. Syst.*, 2013, pp. 1–12.
[64] K. C. Williams, R. L. Spiro, and L. M. Fine, "The customer-salesperson dyad: An interaction/communication model and review," *J. Pers. Selling Sales Manage.*, vol. 10, no. 3, pp. 29–43, 1990.
[65] L. Werner, B. Hanks, H. McDowell, H. Bullock, and J. Fernald, "Want to increase retention of your female students?" *Comput. Res. News*, vol. 17, no. 2, 2005. [Online]. Available: https://users.soe.ucsc.edu/~charlie/projects/pairprogramming/Pipeline.pdf
[66] G. A. Hornstein, "Intimacy in conversational style as a function of the degree of closeness between members of a dyad," *J. Pers. Social Psychol.*, vol. 49, no. 3, pp. 671–681, 1985.
[67] A. B. Kelly, F. D. Fincham, and S. R. H. Beach, "Communication skills in couples: A review and discussion of emerging perspectives," in *Handbook Communication Social Interaction Skills*, J.O. Greene, B.R. Burleson, Eds. Mahwah, NJ, USA: Erlbaum, 2003, pp. 723–752.
[68] R. M. Krauss and E. Morsella, "Communication and conflict," in *The Handbook of Conflict Resolution*, M. Deutsch, P. T. Coleman, E. C. Marcus, Eds. 2nd ed. San Francisco, CA, USA: Jossey-Bass, 2006, pp. 144–157.
[69] R. Rosenthal and R. L. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis*, 3rd ed., Boston, MA, USA: McGraw-Hill, 2008.
[70] E. U. Okike and O. A. Amoo, "Problem solving and decision making: Consideration of individual differences in computer programming skills using Myers Briggs type indicator and Chidamber and Kemerer Java metrics," *J. Appl. Inf. Sci. Technol.*, vol. 7, no. 1, pp. 27–34, 2014.
[71] E. Yourdon, *Death March*. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.
[72] J. E. Hannay, E. Arisholm, H. Engvik, and D. I. K. Sjøberg, "Effects of personality on pair programming," *IEEE Trans. Softw. Eng.*, vol. 36, no. 1, pp. 61–80, Jan./Feb. 2010.
[73] L. Murphy, B. Richards, R. McCauley, B. B. Morrison, S. Westbrook, and T. Fossum, "Women catch up: Gender differences in learning programming concepts," *ACM Special Int. Group Comput. Sci. Educ. Bull.*, vol. 38, no. 1, pp. 17–21, 2006.
[74] K. Stapel, E. Knauss, K. Schneider, and M. Becker, "Towards understanding communication structure in pair programming," International Conference on Agile Software Development XP 2010: Agile Processes in Software Engineering and Extreme Programming. Trondheim, Norway, pp. 117–131.
[75] C. E. Runde and T. A. Flanagan, *Developing Your Conflict Competence: A Hands-On Guide For Leaders, Managers, Facilitators, and Teams*, 3rd ed. San Francisco, CA, USA: Jossey-Bass, 2010.

Stephen Choi received the B.A. degree in chemistry & economics from Rutgers University New Brunswick, NJ, USA, in 1992 and the B.S. degree in management information systems, the M.S. degree in management information systems, and Ph.D. degree in information systems from New Jersey Institute of Technology (NJIT), Newark, NJ, USA, in 1997, 1998, and 2004. He is an Associate Professor with the Information Systems and Decision Sciences Department, Craig School of Business, California State University Fresno, Fresno, CA, USA. His research interests include human-oriented information systems, social and mobile computing, and artificial intelligence. Prof. Choi has authored or coauthored journals in the *International Journal of Electronic Commerce*, *Computers in Human Behavior*, *Journal of Information & Software Technology*, and the IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION.

Research Article

Privacy Rating: A User-Centered Approach for Visualizing Data Handling Practices of Online Services

—Susanne Barth, Dan Ionita, Menno D. T. de Jong, Pieter H. Hartel, and Marianne Junger

***Abstract*—Background:** *Many countries mandate transparency and consent when personal data are handled by online services. However, most users do not read privacy policies or cannot understand them. An important challenge for technical communicators is empowering users to manage their online privacy responsibly.* **Literature review:** *Research suggests that privacy visualizations may alleviate this problem, but existing approaches are incomplete and under-researched.* **Research questions:** *1. How can we design a privacy rating that optimally empowers users with different levels of knowledge about and awareness of online privacy? 2. How do users react to such a privacy rating, in terms of usability, perceived usefulness, and trust in online services?* **Methodology:** *We developed Privacy Rating, a tool for mapping and visualizing the privacy of online services. The tool was subjected to user research (N = 30) focusing on usability, perceived usefulness, and effects on trust. To establish the effects on trust, participants were exposed to a website with either a positive or a negative privacy rating.* **Results:** *The Privacy Rating appeared to be usable and useful for lay users, and it had a significant effect on users' trust in the online service. Users indicated that they would like the visualization to become an established standard, preferably approved by an independent organization.* **Conclusions:** *The Privacy Rating is a user-friendly privacy visualization covering all relevant aspects of privacy. We aim to bring the tool to the market and make it a standard, ideally supported by an independent trustworthy organization.*

Index Terms*—Online privacy, privacy rating, privacy visualization, usability, user-centered design.*

Imagine giving a complete stranger your address and phone number, the contact information of everyone you know, unlimited access to your photos, a detailed account of your media use, all your private messages, and real-time updates on your whereabouts. It sounds extreme, but most of us risk doing just that every day—simply by using online services. Online services ranging from social media and entertainment to shopping and banking continuously handle large amounts of our personal information. The pervasiveness of digital media in modern life has resulted in a semantic web built almost entirely on personal data [1].

Using online services inevitably requires making decisions about disclosing personal data. Disclosures may have adverse consequences such as misuse, spam, or identity theft [2]–[4]. However, due to the complex, multifaceted, and intangible nature of online privacy, the vast majority of users have difficulty judging potential privacy risks and safeguarding their privacy [5], [6]. Privacy policies detail how online services handle user data, but because they are long and complex, few users try to read them, and those few face difficulties understanding them [7]–[11]. Furthermore, an analysis of privacy statements showed that such disclaimers often place little emphasis on providing users with clear-cut information designed to aid the decision-making process. In fact, self-interest and the desire to avoid litigation have much higher priorities among most online service providers [12].

The complex, multifaceted, and intangible nature of online privacy may amplify the cognitive biases that users already have, including optimism bias (underestimating the risks of unsafe behaviors), status quo bias (exhibiting an affinity for default choices), app desirability bias (adjusting privacy concerns based on the attractiveness of the app), and anchoring (taking other users' behaviors as a reference point) [13], [14]. A recent study showed that, in line with Festinger's cognitive dissonance theory [15], users tend to consider privacy less important when they think that they are not in control [6].

Online privacy does not occupy a prominent position on the research agenda in technical and professional communication, with very few research articles in the last 15 years devoted to the

Manuscript received January 9, 2021; revised April 29, 2021; accepted April 30, 2021. Date of publication November 11, 2021; date of current version November 24, 2021.
(Corresponding author: Menno D. T. de Jong.)
The authors are with the University of Twente, 7500 AE Enschede, the Netherlands (email: s.barth@utwente.nl; d.ionita@utwente.nl; m.d.t.dejong@utwente.nl; pieter.hartel@utwente.nl; m.junger@utwente.nl).

IEEE 10.1109/TPC.2021.3110617

Practitioner Takeaway

- This study describes the design and evaluation of a privacy visualization aimed at empowering users to manage and protect their online privacy responsibly.
- Functional complexity is a major design challenge: Empowering users implies making them aware of privacy risks and giving them shortcuts, as well as access to more detailed information in a clear, concise, and intuitive design.
- User research shows that the Privacy Rating fulfills the needs of users: It is usable and useful and significantly affects users' trust in online services.
- A mechanism for an objective third party to evaluate or certify the privacy visualization for online services is highly desirable.

topic [16], [17], none of which address the challenge of empowering users to act in accordance with their own privacy interests. We believe that online privacy deserves more attention within our discipline because it is an increasingly prominent and inherently complex aspect of the interaction between humans and technology, and could benefit from the verbal and visual communication competencies that typically define the strength of our discipline.

When it comes to empowering users to assume informed responsibility for their online privacy, many researchers have drawn attention to the potential of using privacy labels, visually depicting the threats to privacy associated with online services [18]–[34]. In fact, the European General Data Protection Regulation (GDPR) mandates standardized icons to provide an overview of the intended data processing [35]. In this article, we describe the development and evaluation of the Privacy Rating, a new privacy visualization that we have developed for online services. The label is the result of a research-based inventory of important privacy risks. It includes an efficient tool for mapping privacy features and is designed to raise privacy awareness among nonengaged users and to provide relevant, well-organized information to users who are already concerned about privacy. After a literature review, we describe the privacy label and its rationale before reporting on the design and results of a user test that focused on its usability, perceived usefulness, and effects on users' trust in an online service.

Literature Review

Why is There a Need to Visualize Privacy?

Although users claim to care about their online privacy and have concerns about privacy violations, they generally do not behave accordingly. They download apps, give permissions, and provide personal information without much thought about the potential ramifications of their actions. This discrepancy between attitude and behavior is known as the "privacy paradox" [36]. Research shows that there may be three underlying mechanisms.

1. Users rationally weigh the benefits of downloading an app, giving permissions, or providing personal information against the associated privacy risks.
2. Users have trouble weighing costs and benefits, and instead rely on (possibly biased) heuristics or cognitive shortcuts.
3. Users do not even consider the privacy aspects of downloading an app, giving permissions, or providing certain information [37].

The distinction between these mechanisms may not always be clear in practice. Through their behavior, users put themselves at unnecessary risk. The current situation is a vicious cycle. Virtually all privacy policies are complex and "take-it-or-leave-it." Therefore, individual users have no real choice but to accept online services on their (unclear) terms, a situation that panders to the strategies of many service providers. Although online privacy is a topic of vivid discussions in the academic literature, in practice, it is often reduced to momentary feelings of unease and uncertainty in users.

Designers and providers of online services are in the best position to make data handling processes more transparent to users. Since the end of the last century—even before the introduction of smartphones—researchers have advocated for and worked on a Privacy-by-Design paradigm [38], [39]. Its basic premise is that privacy should be incorporated into the fabric of online services instead of "bolting it on" after the fact. Many Privacy-by-Design standards and guidelines have emerged, (e.g., ISO/IEC 29100:2011) [40].

Although this approach can make a tremendous contribution to users' online privacy, several authors have warned of legal and practical complications [41], [42], as well as problems of adoption and implementation [43]–[45].

In practice, many providers of online services still try to discourage users from exercising their rights to privacy [46]. In addition, a core characteristic of online services is personalization which, by definition, involves some degree of personal data processing. Research shows that different users may have different tolerances of specific data handling practices [47].

Another solution would be empowering users to consciously take more responsibility for their online privacy. This could entail increased education: providing users with more knowledge about the business models of online services, the potential privacy risks of transactions, the exact meanings of permissions, and the best protection methods. However, research suggests that general knowledge and privacy awareness play no significant role in the privacy paradox: Advanced computer science students and even privacy and security experts appear to struggle with the same issues as lay users, exhibiting similarly unsafe behaviors [48], [49].

From a document design perspective, there may be a lot to gain from better information about privacy risks. Given the shortcomings of current privacy statements [7]–[12], some researchers investigated whether or not textual improvements could help. An experimental study showed that merely simplifying privacy statements based on document design principles does not affect users' comprehension, attitudes, or behavior [50]. On the other hand, another experimental study showed that concise and simple privacy warnings do have an effect on users' risk perceptions and online behavior [51].

Beyond their legal jargon and complexity at the word-, sentence-, and paragraph-level—all severe problems in their own right—privacy statements generally represent an intimidating information overload that does little to align with the perspective of users trying to ascertain whether to use an online service or not. It seems important to realize that there is functional complexity involved when communicating privacy risks [52], [53]. Ideally the same privacy information should do the following:

1. Raise users' awareness of the importance of privacy and privacy risks [54], [55]
2. Provide less engaged users with a shortcut to support their decision-making about the potential privacy risks associated with using an online service
3. Provide highly engaged users with user-friendly and comparable information about privacy risks (with varying levels of detail, depending on their interests and expertise)

Privacy visualizations, as advocated and developed by several researchers [18]–[34], may be a viable way to address this communication challenge. More than verbal information, visualizations can draw the attention of users who are not aware of privacy risks [24], [33], [56] and force service providers to translate complex privacy information into manageable, standardized privacy information.

Earlier Attempts to Visualize Online Privacy

Developing a privacy visualization requires two related activities: an intrinsic analysis of the relevant privacy aspects to be included and a verbal-visual communication design. Both in the academic literature and in practice, many attempts have been made to develop privacy visualizations (see Barth et al. [57] for an overview). Table I summarizes 14 earlier attempts, with special attention to the extent to which the systems provide overall advice about the privacy risks of online services (overall indicator) and detailed information about specific privacy aspects (privacy details).

Existing privacy visualizations operationalize privacy information quite differently [34], [57]. Barth et al. [57] investigated operationalizations of online privacy that manifest themselves in Privacy-by-Design guidelines and privacy visualizations, resulting in the following 15 different privacy aspects:

- Accountability
- Anonymization
- Collection
- Control
- Correctness
- Disclosure
- Functionality
- Purpose
- Pseudonymization
- Retention
- Right to be forgotten
- Sale
- Security
- Sharing
- Transparency

TABLE I
OVERVIEW OF EARLIER PRIVACY VISUALIZATIONS

Year	Source	Name	Type	Overall Indicator	Privacy Details
2007	[58]	Mehldau's data privacy declarations	Icons with tags	No	Yes
2009	[18], [19]	CyLab's privacy nutrition label	Table	No	Yes
2009	[59]	KnowPrivacy's policy coding	Icons with tags	No	Yes
2010	[60]	Mozilla's privacy icons	Icons	No	Yes
2010	[20], [21], [61]	PrimeLife privacy icons	Icons	No	Yes
2011	[62]	TrustArc's privacy short notice	Icons	No	Yes
2012	[22]	Privacy wheel	Privacy label	Yes	Yes
2014	[23], [25]	GDPR's draft privacy icons	Icons	No	Yes
2017	[28]	Data controller indicators	Data flow representation	No	Yes
2018	[29]	Renaud and Shepherd's privacy summary	Summary supported by icons	No	Yes
2018	[31]	Fox et al.'s GDPR compliant privacy label	Summary supported by icons	No	Yes
2019	[26], [27]	Data protection icon set (DaPIS)	Icons with tags	No	Yes
2019	[63]	Clever°Franke's privacy label	Privacy label	Yes	Yes
2020	[64]	Privacy label	Summary supported by icons	No	Yes

None of these privacy aspects were incorporated in each of the reviewed privacy visualizations. Three privacy aspects were quite prominent—types of data collection, purposes of data collection, and data sharing—with only one or two visualizations missing out on them. But the overall focus of the visualizations differed considerably. No agreed-upon framework of relevant privacy aspects of online services currently exists. A new privacy visualization should thus be based on a systematic analysis of relevant aspects of online privacy.

Various types of visualizations can be distinguished. Seven of the 14 visualizations listed in Table I are sets of icons expressing specific privacy characteristics. Several authors have argued that it is difficult to visualize such intangible and complex features [22], [25], [65], and several icons that were developed proved to be problematic in user tests [20], [61]. As a result, some of the icon sets use supporting tags to assist with the interpretation of visual cues. A significant drawback of icons is that they are limited to depicting specific privacy risks, thus making them unsuitable for providing users with the bigger picture, which is necessary if they are to make informed decisions about the acceptability of the combined privacy risks.

Three of the proposed visualizations downplay the role of icons by making them merely supportive for predominantly written information. In these cases, the icons have no independent meaning but only visually support the structure of a summarized privacy text. Again, it is questionable whether this approach supports users in their decisions about the combined privacy risks of online services. The difficult task of making sense of the various privacy characteristics and translating those insights into an overall judgment about privacy risks is still entirely the users' responsibility.

Two other visualizations explore very different directions. Inspired by the nutrition labels on food, Kelley and colleagues developed a privacy nutrition table, which actually consists of a listing of 10 types

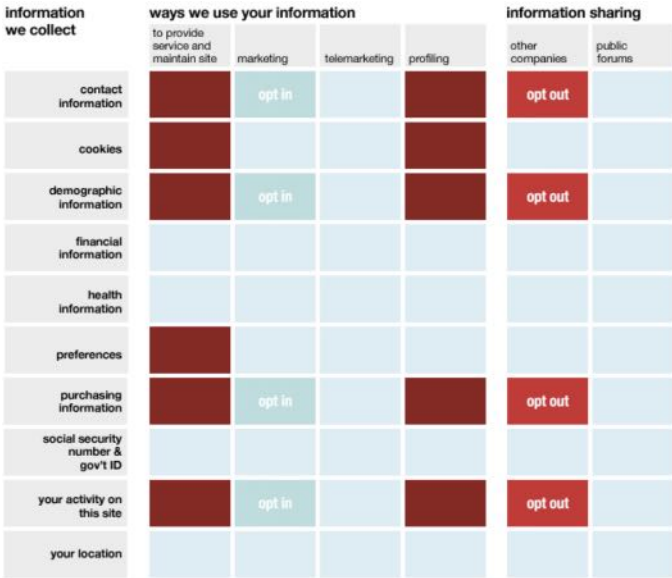


Fig. 1. Kelley et al.'s privacy nutrition label [18], [19].

Fig. 2. Van den Berg and Van der Hof's privacy wheel [22].

of user data, five types of data handling, and two different parties handling the data (see Fig. 1). In each cell of the table, four options may be entered (yes, no, opt out, and opt in) [18], [19]. The analogy with nutrition labels already suggests that the visualization does not attract less-engaged users and does not support users' overall decisions about whether privacy risks are acceptable or not. Still, a focus group study showed that users appreciated the system [18], and a comprehensive experiment showed that the label, compared to normal privacy statements, helped users to better understand the privacy aspects of online services [19].

Van Kleek et al. developed a visualization of the data flows from online services [28]. Although the resulting graphs were advanced and may be too complex to be intuitively comprehensible, a small-scale experimental study indicated that the visualization, more than written privacy information, helped users make informed decisions regarding online privacy.

Finally, two proposals for visualizations take the form of privacy ratings, providing overall indications of the privacy aspects of online services with optional in-depth information. Van den Berg and van der Hof's privacy wheel (see Fig. 2) consists of an overall privacy qualification in the middle surrounded by eight brightly colored clickable aspects of privacy [22]. Although it manages to combine an overall privacy assessment and more detailed information, the visualization has a few potential drawbacks: The overall privacy assessment in the middle might be easily overlooked, it lacks a reference point, and it is not transparently related to the eight specific privacy aspects.

Clever Franke's privacy label (see Fig. 3) is inspired by the letter classification (A-F) and color use of the EU energy label [63]. It consists of a colored circle with a privacy qualification in the middle: An A (in green) is positive; an F (in red) is negative. Around the qualification, there is a circle divided into three equal parts representing three privacy aspects: data usage, data collection, and user control. For every aspect, five questions are asked. For positive answers, the line is colored; for negative answers, it is left white. The thicker the colored circle around the privacy qualification, the more positive the online service scores on the specific privacy aspects. Users can use a quick response (QR) code for more specific information. Drawbacks of this visualization are that the specific privacy information is hidden in the design, and the system of five questions in three parts of the circle may not be clear to users.

No research reports are available on user tests with either of these two privacy labels.

Research Questions

In this article, we describe a project developing a Privacy Rating tool for online services that is

Fig. 3. Clever Franke's privacy label [63].

founded upon expert knowledge of the relevant privacy aspects and that is designed to overcome the shortcomings of earlier privacy visualizations. Furthermore, we describe a user study of the proposed visualization that focuses on usability, perceived usefulness, and effects on user trust. We investigate the following research questions.

RQ1. How can we design a Privacy Rating tool that optimally empowers users with different levels of knowledge about and awareness of online privacy?

RQ2. How do users react to such a Privacy Rating tool, in terms of usability, perceived usefulness, and trust in online services?

Methodology

A user-centered privacy visualization must be both useful (contain the right information) and usable (present the information in an understandable way). Therefore, our methodology is two-fold. First, we report on the development of the Privacy Rating visualization. At the core of the proposed visualization lies a set of 12 privacy metrics that not only provide the basic structure and content of the privacy visualization, but also serve as input for the rating system. Second, as a part of an iterative design process, we evaluated the visualization with potential users.

Privacy Rating Below, we describe the development of the Privacy Rating visualization and discuss its three main characteristics: content, visual design, and generating the Privacy Rating.

Content: The development of the Privacy Rating started with a thorough and systematic analysis of the privacy aspects of online services that should be deemed relevant and therefore included. We took the list of 15 privacy attributes gathered in earlier research [57] as our starting point (see Table II). The attributes were based on established Privacy-by-Design guidelines and earlier privacy visualizations. Research with experts and users confirmed the importance of all attributes [57].

We decided to exclude two of the original attributes for our visualization. The functionality aspect was removed because it was ambiguous and overlapped with control. Transparency was removed because having a Privacy Rating can already be seen as a positive indicator of transparency in itself. In addition, anonymization and pseudonymization were combined into one attribute because they were sometimes difficult to distinguish: pseudonymization can be seen as incomplete anonymization. From previous research, we know that privacy is subjective and context-dependent [47], [66], [67]. Therefore, we decided to use all of the remaining 12 attributes as equally rated metrics for our rating system.

Because differentiating 12 different privacy attributes is not manageable for users, we conducted a card-sort study in which we asked users to cluster the 12 attributes. Most often, the attributes were grouped into four categories. Although security turned out to be a clear group label, there was disagreement about the others. Consulting 10 privacy and cyber security experts from our network resulted in four main clusters: collection, sharing, control, and security (see Table III).

To use the metrics for rating and comparing online services, they must be operationalized. To keep the system simple and understandable for users, we defined three-point scales (good-neutral-bad) for each attribute. In iterative sessions with privacy and cyber-security experts, we arrived at the operationalized metrics presented in Table III. Online services receive penalty points depending on their score on each metric (0 points for good scores, 1 point for neutral scores, and 2 points for bad scores).

TABLE II
PRIVACY ASPECTS CONSIDERED FOR THE PRIVACY RATING [57]

Privacy Aspect	Description
Accountability	Can the service provider be held accountable for violations?
Anonymization**	Are all identifiable markers completely removed so that data can never be traced back to individual users?
Collection	Which user data are collected?
Control	Are users able to choose or decide which data to share for which purpose, and how difficult is it to do so?
Correctness	Are there mechanisms for preventing and fixing incorrect data?
Disclosure	What is the provider's attitude towards data requests from law enforcement?
Functionality*	Are users forced to choose between functionality and privacy?
Pseudonymization**	Are personally identifiable markers replaced by artificial identifiers, or pseudonyms, so that data can only be traced back to individual users with the help of additional information?
Purpose	What are the collected data used for?
Retention	How long are collected data stored?
Right to be forgotten	Can users request that all their personal data will be removed?
Sale	Are any of the data sold to third parties?
Security	Which technical measures are taken to ensure that data are protected from unauthorized or malicious access?
Sharing	Do any of the collected data leave the ownership of the provider?
Transparency*	Are users able to obtain information about how their personal data are handled?

Note: $*$ = Removed from the *Privacy Rating* attributes; $**$ = Collapsed into one attribute.

Fig. 4. Design of Privacy Rating.

The total number of penalty points is then used to categorize online services into seven classes, from A (lowest privacy risks) to G (highest privacy risks):

- Class A: 0 or 1 points
- Class B: 2 to 5 points
- Class C: 6 to 9 points
- Class D: 10 to 13 points
- Class E: 14 to 17 points
- Class F: 18 to 21 points
- Class G: 22 to 24 points

Visual Design: Our Privacy Rating (see Fig. 4) was designed through an iterative process in collaboration with a professional design agency. Simplicity, clarity, recognizability, and attractiveness were important criteria throughout the design process. With its stable and marked overall design, the visualization has the potential to draw attention to privacy issues across different online services. The use of overall privacy classes helps less-engaged users to make a quick overall

TABLE III
CLUSTERED AND OPERATIONALIZED PRIVACY ATTRIBUTES

Cluster	Attribute	Operationalization
Collection	Collection	0 - Collects anonymous data
		1 - Collects personal data, relating to an identified or identifiable person
		2 - Collects sensitive data, involving racial or ethnic origin, political views, religious or philosophical beliefs, trade union membership, genetic or biometric data, health status, or sexuality and sexual orientation
	Purpose	0 - Used for functionality only
		1 - Used for customization (personalization in the current interaction)
		2 - Used for profiling
	Retention	0 - Data not stored
		1 - Data stored for a pre-determined limited time
		2 - Data stored indefinitely
Sharing	Sharing	0 - No sharing of user data
		1 - Sharing of anonymous user data
		2 - Sharing of user data
	Sale	0 - No sale of user data
		1 - Sale of anonymous user data
		2 - Sale of user data
	Disclosure	0 - Statutory disclosure to local law enforcement (inside user's jurisdiction)
		1 - Disclosure to local law enforcement (outside user's jurisdiction)
		2 - Disclosure to foreign law enforcement
Control	Control	0 - Opt-in (users must explicitly opt-in to allow data collection)
		1 - Opt-out (data are collected by default, but users can opt-out)
		2 - No opt-in or opt-out
	Right to be forgotten	0 - Data deleted upon request
		1 - Data hidden upon request
		2 - Data cannot be removed
	Correctness	0 - All data can be amended
		1 - Some data can be amended
		2 - Data cannot be amended
Security	Security	0 - Industry standard security (certified compliant with the latest version of either ISO 27001 or NIST 800-53)
		1 - Basic security (developed in compliance with the OWASP Top 10 standard and tested according to the OWASP Application Security Verification standard or the OWASP Mobile/Web Security Testing Guide or equivalent)
		2 - None of the above
	Anonymization	0 - Anonymous (all identifiable markers are completely removed so that collected data can never be traced back to individuals)
		1 - Partially anonymous (personally identifiable information fields within collected data are replaced by artificial identifiers or pseudonyms, so that data can only be traced back to individuals with additional information)
		2 - Not anonymous (personally identifiable information is stored)
	Accountability	0 - Legally accountable
		1 - Legally binding privacy policy
		2 - Not legally accountable

Note: 0-2 represents the number of penalty points for each alternative.

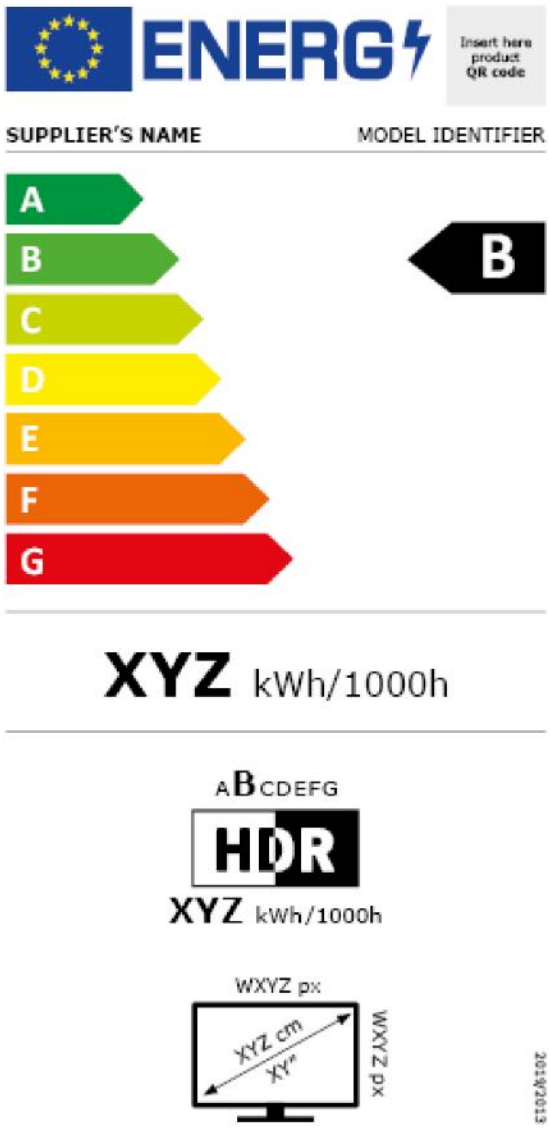


Fig. 5. Template of the European energy label for electronic displays [68], [69].

judgment about the potential privacy threats of online services. As with the familiar European Union energy label (see Fig. 5) [68], [69], privacy classes are indicated by combinations of letters and colors (ranging from A plus green for the most positive online services; to G plus red for the most negative ones). The colors also reflect the conventional color scheme of traffic lights. The presence of a full scale helps users to interpret the score of a particular online service.

Users who are more engaged with online privacy are helped with two levels of additional specific information. The first level, immediately obvious in the visualization, is the scores of the online service in the four main categories of privacy aspects (collection, sharing, control, and security), which can have different colors depending on the specific score for each one. Each category is listed with its name and an icon. The second level, which can be reached by hovering over or clicking the main categories, provides more detailed information about specific aspects of privacy.

Generating Privacy Rating: To promote the practical feasibility of the Privacy Rating, we developed a self-assessment form in a free web application (www.privacyrating.info). This form enables providers of online services to create their own tailor-made privacy label corresponding to the data handling practices of their online service. The application is designed to walk service providers through a questionnaire with each question corresponding to one of the three levels of each attribute. The questionnaire is interactive: Once the answer to a question confirms the level of an attribute, the remaining questions corresponding to that attribute are skipped, and the service provider is directed to questions about the next attribute. When all 12 attributes have been evaluated, the application computes the Privacy Rating and creates a visualization in two formats: an HTML and a smaller PNG version, both of which can be embedded into webpages or apps. The small version can be added to the footer of the page or to the cookie notice. The larger version can be included in the privacy policy or as a pop-up.

Research Design of the User Study To evaluate the potential value of the Privacy Rating, we conducted a user study. In this early phase of development, we focused on the following three aspects of the privacy label:

- Usability
- Perceived usefulness
- Effect on users' trust in an online service

Because of the COVID-19 pandemic, data collection took place in individual online sessions. The study was approved by the Ethical Committee of the Electrical Engineering, Mathematics and Computer Science faculty of the University of Twente.

Participants: Participants were recruited in three complementary ways:

- From the university's research participants pool
- From a commercial research participants pool
- Via social media

Participants from the university's pool received participant credits required by their study

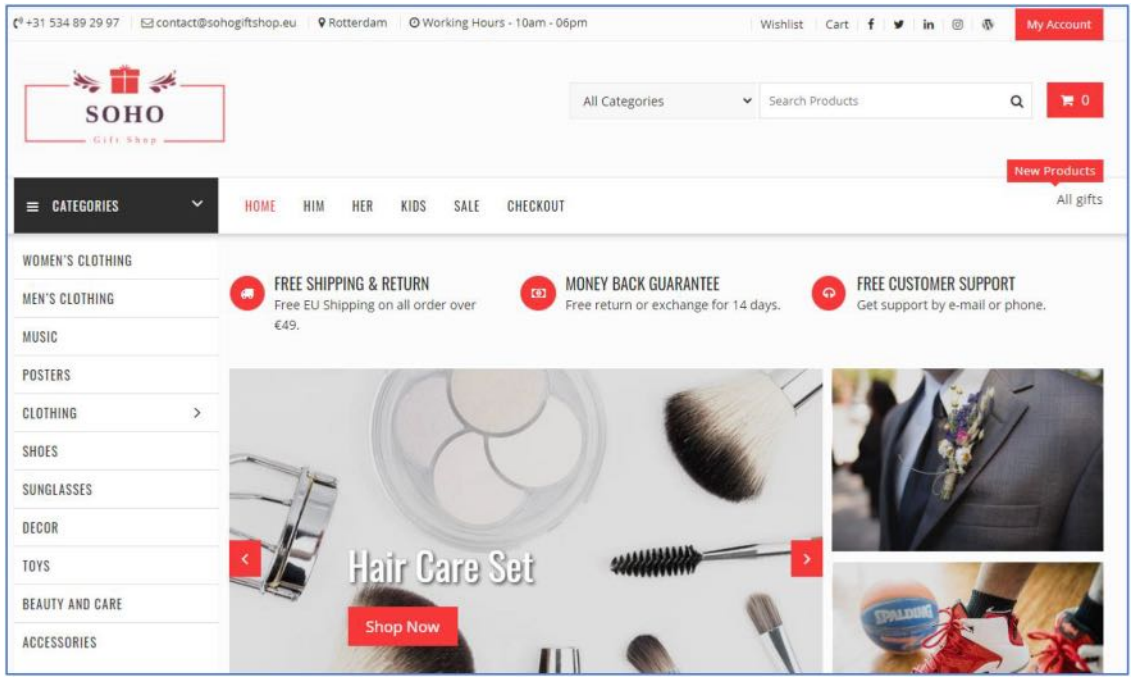


Fig. 6. Screenshot of the web shop for the user study.

programs, participants from the commercial pool received a monetary compensation, and participants from social media volunteered to participate without compensation. In our recruitment messages, we called for participants aged 18 or older, with good English proficiency and access to a Google Chrome browser, a web cam, and a microphone.

A total of 30 participants took part in the study. Participants had a mean age of 28.6 years (ranging from 19 to 62). Their gender distribution was equal. Participants' educational level varied from medium (high school or vocational education: 53%) to high (bachelor, master, and PhD: 47%). Of the sample, 60% currently followed a study program, and 57% had a job. Study programs and occupations were quite diverse. Three participants had a background in cyber security or online privacy. All participants lived in Europe, most of them coming from Germany or the Netherlands. A large majority of the participants had ample experience with online tools such as email, search engines, instant messaging, social media, and teleconferencing (all 93% or higher) and with online transactions such as online banking, streaming, and shopping (all 87% or higher).

Research Materials: To evaluate the Privacy Rating in a realistic setting, we built an online web shop (see Fig. 6), using a real, SSL-protected domain (www.sohogiftshop.eu). The web shop used a prebuilt, highly rated WordPress theme. Offerings (including photos, descriptions, and prices) were selected across a broad range of product types. To prevent unintended visitors, the web shop was password protected. Participants received the password at the beginning of their session.

The web shop's Privacy Rating was included as a pop-up that appeared when users opened the homepage. Before interacting with the site, users had to click away the pop-up. The shop's Privacy Rating was also included at the bottom of the homepage, and a small version was added to the footer of every page (see Fig. 7).

To investigate the effects on participants' trust in an online service, two versions of the Privacy Rating were used: Half of the participants were exposed to the web shop with a moderately positive Privacy Rating (grade B, predominantly green), and the other half were exposed to the web shop with a moderately negative rating (grade F, predominantly red).

Procedure: The research sessions consisted of two parts. Participants began with an online questionnaire in Qualtrics covering their background characteristics and the consent information. Background questions focused on age,

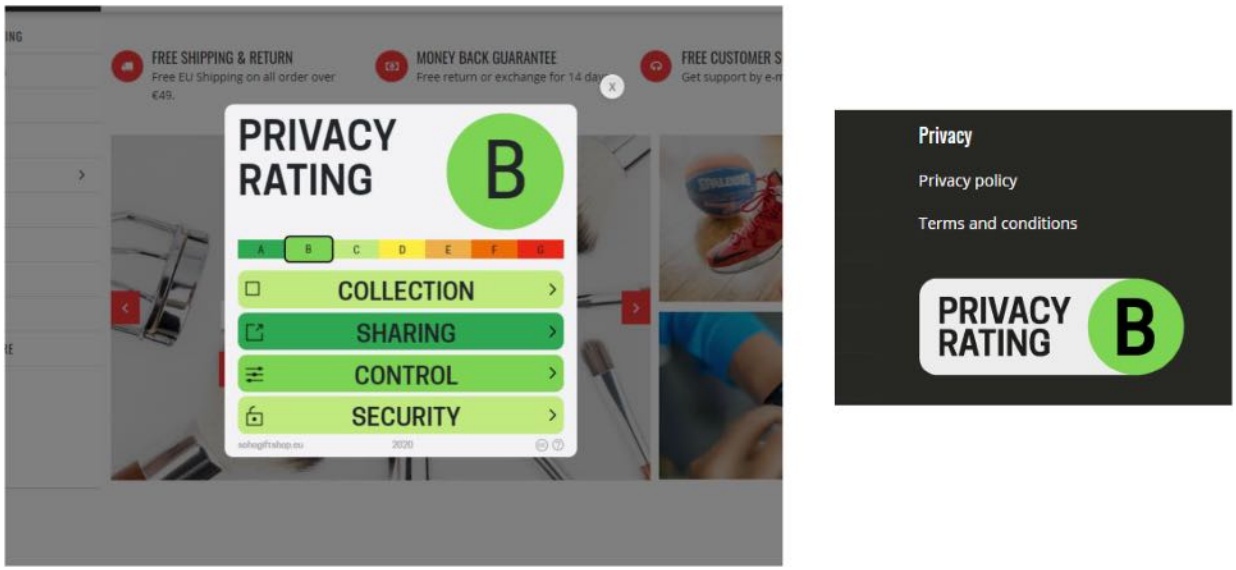


Fig. 7. Privacy Rating on the web shop, as pop-up (left) and as small label (right).

gender, country of residence, education, profession, use of online services, and expertise in online privacy and cyber security. After filling out all questions, participants received a link to a live session with one of the interviewers. In the live sessions, we used Lookback for real-time screen monitoring and interviewing. Participants were asked to install this software on their computers. In all sessions, two researchers were involved: one moderated the session and interviewed the participant; the other observed without interacting with the participant.

The session started with the following scenario-based task.

> You are looking for a gift for the birthday of your friend. You find some interesting gifts in an online gift shop you have never used before: the SOHO Gift Shop. To place an order, you must provide your first and last name, date of birth, gender, age, shipping and billing address, and credit card details. Try to determine whether you would trust this website with your personal information.

To avoid reactivity, we did not ask the participants to think aloud. However, their interactions with the Privacy Rating and the website were recorded and used in the analysis.

The task execution was followed by a semistructured interview, with questions covering the following three topics.

1. Trust in the website
   a. Decision whether to make a purchase
   b. Impression of the website
   c. First impression of the Privacy Rating
   d. Effects of the Privacy Rating on trust
2. Usability of the Privacy Rating:
   a. Name
   b. Overall rating
   c. Scale
   d. Main categories
   e. Detailed information about the categories
   f. Visual design
3. Usefulness of the Privacy Rating:
   a. Transparency (did it increase an understanding of data handling practices?)
   b. Behavioral intentions (would it affect decisions to trust online services?)
   c. Desirability (would the participant like to see it as an established standard?)

The sessions were videorecorded. Sessions lasted on average 24.4 minutes (SD = 8.3). At the end of the sessions, participants were thanked, debriefed, and given instructions for removing the Lookback extension from their browser.

Analysis All 30 interviews were transcribed verbatim, and any personal information that could be associated with participants was removed. The interview data were analyzed qualitatively in ATLAS.ti. Codes were based on the interview questions and emerged bottom-up based on participants' answers. Two independent researchers coded a random selection of 10% of the

transcripts and discussed the discrepancies in their coding. Based on the discussion, the coding scheme was refined. After that, the two researchers coded another sample of the transcripts. They reached sufficient intercoder agreement in general (Cohen's kappa = 0.85) and for the three main research topics: usability (0.87), perceived usefulness (1.0), and trust (0.78). Using this coding scheme, the remaining transcripts were then coded by the first author.

To investigate the effects of the Privacy Rating on participants' trust in the online service, the interviews were complemented with behavioral data: the amount of time participants spent looking at the Privacy Rating pop-up and their decision about placing an order in the web shop. For these behavioral data, we compared the results of the two experimental groups (positive versus negative Privacy Label).

Results

Usability

Name: Most participants (80%) found the name Privacy Rating clear and understandable and formulated correct expectations of its purpose: "It's really clear that this is about how safe a website is in terms of privacy." Others stated that they would not know immediately what the name "is trying to communicate." To come to a full understanding, they would have to see more. Interpreting the name in combination with the other elements helped them to "understand what they mean, what they tell you."

Overall Rating: Participants were generally positive (87%) about the clarity of the overall rating: "It's understandable enough to make me not want to share my information." For most participants, the color was important: "If there would be no color, it could be like, what does B mean? But green is always good and red is bad." Some participants related the overall rating to other familiar grading or rating systems: "the labels for energy consumption," "the American paper grading system," or "the alphabet; where the alphabet starts, the better it is." Participants with difficulties understanding the overall rating stated that the meaning became clearer when they also looked at other elements (for example, the colored scale).

Scale: Most participants (80%) found the scale clear: "A would mean that this is the best rating of privacy that you could have as a website, and G would be the worst." Some called the scale "intuitive" and "nothing to misunderstand." The use of colors makes it easy to interpret.

> A is green. So like a traffic light, green is good. Green, you go, you're safe to go. Yellow as well, you can go. … And then red is no, you don't go. Not very good.

Some participants stated that a scale without colors would be harder to understand. Others said that they needed a point of reference to interpret the scale. Interestingly, two participants expected that the scale would be interactive with clickable letters.

Main Categories: The four main categories (collection, sharing, control, and security) were clear to most of the participants, although some argued that the terms alone did not suffice and were understandable only when looking at the details corresponding to the categories. The categories collection and sharing were easiest to understand (93% and 87%, respectively). Control and security were somewhat less clear to the participants. Regarding control, several participants (77%) found the term "a bit vague" and "difficult to understand." Some thought it referred to the control service providers have—"maybe what the website can do remotely to your computer"—and did not see that it is meant to refer to the control users have regarding their personal data. Regarding security, participants (70%) found the term "a bit ambiguous" or "too general." Some thought that it involved only financial transactions: "Should I give my Visa number or should I use PayPal?"

Detailed Information: Most participants (63%) found the more detailed information underlying the four categories clear. Although participants appreciated the conciseness of the descriptions, some suggested adding more information because it "is very much open to interpretation depending on how much knowledge the individual has." Some participants found the wording too technical and would have appreciated explanations in "more human [layman's] words." Table IV summarizes the specific problems participants mentioned about the detailed information.

Visual Design Elements: Three participants found the separate colors used for the four categories confusing. One participant found the green color difficult to see against its background. Another

TABLE IV
PROBLEMS IDENTIFIED IN THE DETAILED INFORMATION

Collection	Sharing	Control	Security
What does "functionality" mean? (n = 4)	What does "legally required disclosure to local law enforcement" mean? (n = 5)	What does "opt-out" mean? (n = 7)	What does 'basic security' mean? (n = 2)
How long is stored for a limited time? (n = 3)	Is not sharing data realistic? (n = 1)	What does "amended" mean? (n = 2)	What does "industry-standard security" mean? (n = 2)
Which personal data does it collect? (n = 2)			"No anonymization" is a vague term (n = 1)
What does "data stored indefinitely" mean? (n = 1)			What does "legally accountable" mean? (n = 1)

participant understood this color scheme differently, stating that "sharing and using data is red. So I'm assuming that means that they don't share my data," whereas the color red actually means the opposite. In addition, three participants were confused that the categories expanded both automatically (when hovered over) and manually (when clicked).

Several participants (33%) did not realize that the indicators were ratings of the single statements. "The color of the overall rating and the color of the subcategories are the same. I did not notice that those are ratings." Another participant thought that the colored dots were "just simple bullet points that don't have any meaning." Especially the green dots were difficult to recognize "because the background is all green and the bullet points are all green."

The icons used to support the meaning of the four categories were correctly understood and appreciated by 44% of the participants. The other participants had difficulties with one or more of the icons. One participant questioned whether or not the icons are really necessary. The interactivity of the icons, intended to catch the user's attention, proved especially confusing for some participants.

> The fact that they move … I can get a little distracted and it makes it look a little less trustworthy to me and not necessarily helping me better understand what it is about.

Another participant assumed that the icons would be clickable and have a personalization function integrated.

In all, the usability evaluation yielded a positive overall impression as well as several suggestions to further optimize the Privacy Rating (see Fig. 8 for an overview). Some of the detailed problems mentioned with specific elements are actually solved when participants consider the complete visualization. However, the results revealed the need for more attention to the wording of categories and detailed information, with an important balance between clarity and conciseness. In addition, the participants mentioned several ambiguities in the visual design that deserve attention.

Perceived Usefulness Overall, participants were very positive about the Privacy Rating, one of them calling it "the most useful tool I've seen." The vast majority of participants (90%) considered the label to be an effective tool for visualizing how online services handle users' personal data.

> I think it's pretty good for a normal homepage because usually it's not so easy to find this information and I don't usually read all of it unless it's a new company.

The similarities to the existing EU energy label appeared to enhance the label's usefulness. "It reminds me a bit of when you buy a fridge and you get the label in terms of the efficiency levels." Some participants explicitly appreciated that the label was the first thing they saw when opening the website. "It gives a pretty clear overview. And it's also nice if I click on the website and it's right there."

Three participants were somewhat more critical, arguing that the information provided by the label only "gives an impression but not a full clear explanation of how this website is handling my data." Their objections involved the conciseness and clarity of the information, as discussed above.

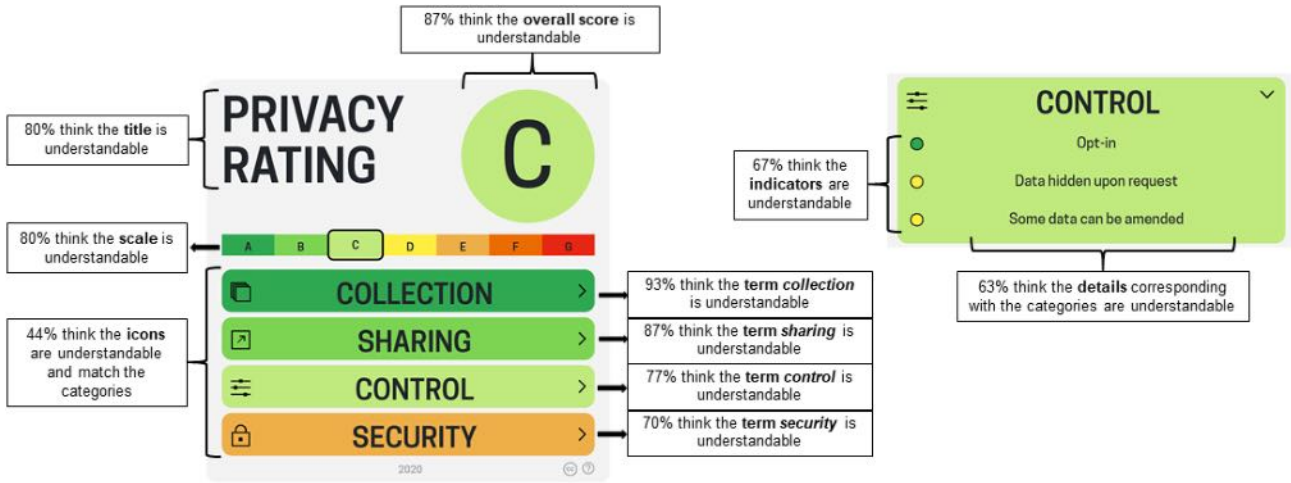


Fig. 8. Usability of the various elements of the Privacy Rating.

Most participants (83%) felt that the Privacy Rating would influence their decisions on trusting and using websites or other online services. They would appreciate such a label, especially when sharing sensitive data such as credit card details with an online service. The label would help them to judge unknown websites or compare services offering the same product. It makes evaluating online services less time-consuming and limits the role of subjectivity in their judgments. Interestingly, some participants argued that a negative rating would influence them more than a positive rating.

All 30 participants would like the Privacy Rating to become an established standard under the responsibility of an independent organization because such standardization would enhance people's awareness of online privacy and the risks of data sharing. It would also educate users, satisfy the needs of users who care about their personal data, and decrease vulnerability to fraud.

> I would be happy to see something like that on a website, generally. It would help educate people as to the good and the bad out of the internet, and shopping online and banking online. I surprised myself … how many online systems I actually use. I worked in IT, but I'd like to think of myself as being able to disconnect from it. But clearly not. Everything I do is connected to technology in some way.

Effects on Trust A first step in our analysis of the effects that the Privacy Rating had on participants' trust in the web shop involved the attention that participants paid to the pop-up. On average,

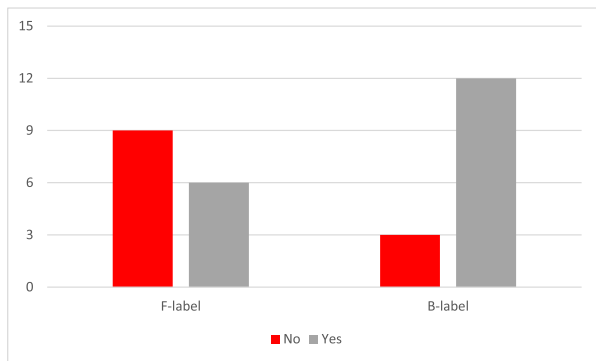


Fig. 9. Effects of the Privacy Rating on participants' decision whether to place an order.

participants spent 33.4 seconds ($SD = 18.5$) looking at the label (with a range between 6 and 78 seconds). There were no significant differences between the groups that had been exposed to a positive or negative label. In the interview afterwards, almost all participants (93%) indicated that they recognized the label; only two participants were not sure whether they had seen it.

A second step was to determine whether the label affected participants' online ordering decisions. A chi-square test showed that this was the case (χ^2 $(1, N = 30) = 5.0$, $p < 0.05$). In the group of participants exposed to the negative privacy rating, only 40% would place an order in the web shop, compared to 80% in the group of participants exposed to the positive rating (see Fig. 9).

Many participants indicated that a negative rating would influence them more than a positive one.

Indeed, participants who saw a negative rating displayed were less likely to place an order compared to those who were shown a positive rating. Furthermore, although a good Privacy Rating increased trust in the website for 66% of the participants, a bad rating decreased trust for 91% of the participants. This finding indicates that, in our sample, a negative rating had a greater influence on trust than a positive rating did.

From the interviews, two possible factors could be identified that might limit the effectiveness of the Privacy Rating. The first is that the pop-up format is not always appreciated. Some of the participants saw it as annoying and disturbing.

> I don't like websites where you have a pop-up straight away. …When I go to a landing page of a website, I want to have a look at the actual website and not deal with pop-ups.

The second is that the label is not yet officially established and therefore unfamiliar. This provoked suspicion among some participants:

> I think it's a bit weird for a website to have that because on other websites that are trustworthy, I don't see it there. … This was a bit unexpected, but unexpected in a negative sense …. it could be that they do this in order to make their website look trustworthy while they're not.

Discussion

Online privacy is an increasingly important issue. Rapid technological developments in information and communication technologies and artificial intelligence have accelerated the impact of computers and mobile phones in our lives as well as the possibilities for online service providers to invade our privacy. Interfaces have become deceptively simple and user-friendly, whereas the processes going on in the background are increasingly complex and opaque. Researchers have spent a lot of time and energy unraveling people's privacy-related attitudes and behaviors, and exploring the privacy paradox, but so far, research-based attempts to empower users with the means to assume responsibility for their online privacy have been limited and unsuccessful.

In this article, we described the design and evaluation of a new privacy visualization called Privacy Rating. To inform users about potential consequences of information disclosure and to raise awareness of data handling practices, a risk-based and multilevel approach was chosen for the design of the Privacy Rating. At the core of the privacy visualization lies a set of 12 privacy aspects derived from a literature review [57]. To avoid information overload, the 12 aspects were divided into four main groups: collection, sharing, control, and security. Dividing otherwise complex information into smaller text passages, in combination with colors and navigation options for more information, makes the information accessible and tailored to individual information needs.

The label acknowledges the functional complexity involved in communicating privacy aspects and supports both less engaged users and privacy-aware users. If widely implemented, it may contribute to privacy awareness among users in general, as it sheds light on the privacy aspects of online services, transforming them from a hidden feature into a conspicuous and comparable characteristic. For less engaged users who may worry about privacy but are unwilling to invest time and effort into evaluating all privacy characteristics, the overall Privacy Rating provides a visual shortcut to support their decision-making process about downloading or using online services. For more engaged users who want to know more about privacy but who may be hesitant to examine the entire privacy policy, the Privacy Rating offers prestructured detailed information in two layers. With these contributions, the Privacy Rating may play a positive role in balancing the unfavorable equilibrium between users and online service providers, in which privacy considerations currently play an inferior role.

The user research that we conducted underlined that the Privacy Rating can be a promising tool to help users safeguard their online privacy and thus limit the privacy paradox [36]. Barth et al. [37] identified three underlying mechanisms of the privacy paradox:

1. A more or less rational weighing of costs and benefits
2. An incomplete and biased weighing of costs and benefits
3. A neglect of privacy considerations

The Privacy Rating should help reduce the influence of the latter two mechanisms. The label urges users to consider privacy aspects in their decisions and reduces biases that they might have when judging privacy risks. As a result, the weighing of costs and benefits will be more systematic and more rational than may currently be the case. That does not mean that the privacy paradox is solved. It is still imaginable that users

could decide in favor of an online service despite privacy risks that they are aware of.

But the discrepancy between attitude and behavior may not be at the core of the problem. People have to make tradeoffs between desires and preferences all the time. The core of the problem is the fact that their decisions are often uninformed. Tackling this deficit is the main purpose of the Privacy Rating tool. The results of our user research suggest that this design is a step in the right direction. With regard to usability, the Privacy Rating did quite well, although participants also uncovered several problems that need to be addressed in future iterations of the label. The problems found mainly concerned the formulation of privacy risks and aspects and details in the visual design. The perceived usefulness was judged very favorably by our users, and the label appeared to significantly affect our participants' decisions on whether to use a particular web shop. User feedback will play a significant role in our future efforts to further optimize the Privacy Rating.

In addition, the results of our user research can be used to inform other privacy visualization projects. Two insights stood out. The first is that connecting a privacy visualization explicitly to users' existing interpretation frames is beneficial. In all parts of our user research, we heard positive remarks about the resemblance of the Privacy Rating to the well-established and familiar energy label, which made our label easy to understand and may have also contributed to the persuasiveness and perceived urgency of the rating. The second is that the development of the label is only half of the story. Several participants in the user research doubted the independence and authoritativeness of the label, letting on that it would make a big difference to them if the label were issued by a trusted source.

Finally, our findings drew attention to two tradeoffs in designing a privacy visualization. The first involves finding a balance between conciseness/simplicity and informativeness. The feedback from some of our participants suggested that they found even the second layer in the information about privacy insufficient. Having said this, we are by no means certain that adding information will make the label better. Our findings lend support to previous work stating that grouping and segmenting information across multiple layers has a positive effect on the understandability of complex information [24] and that color schemes can increase granularity and provide shortcuts for quickly assessing risks [32], [34]. Also, in line with previous work, we found that privacy and security icons have poor understandability [26], [34].

The second tradeoff is between annoying intrusiveness and sheer invisibility. Some of our participants complained about the use of a pop-up, but whether a less intrusive exposure would glean the necessary attention and provide similar effects is questionable. Prior research showed that the timing of users' exposure to privacy notices is very important [70]. The development of any viable privacy visualization must include its effective placement. It is quite possible that the methods of exposure may become less important once the label becomes an established standard [25].

Limitations and Future Work

To our knowledge, this is the first initiative to develop a privacy visualization covering a systematic selection of relevant privacy attributes available in the academic literature, law, and practice. It is also one of the few initiatives to explicitly incorporate user feedback in the process. Still, it is important to keep the following limitations in mind when interpreting the results.

First, the Privacy Rating is still in development. In our user study, we tested a prototype of the privacy label that reflected our knowledge after various studies into user perspectives on online privacy [47]–[49], after a thorough analysis of relevant privacy aspects and earlier privacy visualizations [57], and after an iterative design process including expert and user input. The user study reported in this article provided us with more food for thought, which we will use to further optimize the privacy rating. Specifically, we will look into using simpler language and including links to further information.

Moreover, we foresee three additional developments in the period ahead. We will try to further explore the implementation of the label, which involves gaining support from online service providers, platforms, and legislation. Any advancements may have consequences for users' perceptions of the Privacy Rating. We will also try to make the input for the Privacy Rating more objective and trustworthy. The score that online services currently receive is based on service providers' self-reports in the questionnaire. That is not necessarily a bad option, as service providers can be held responsible for any discrepancy between their privacy policies and their answers in the questionnaire. But ideally, the privacy ratings

would be obtained directly from the privacy policies, either by natural language processing or by the intervention of an independent authority. Future developments in this respect may also have a positive impact on users' perceptions. We will try to set up communication about the Privacy Rating itself. In the current user study, participants saw nothing but the visualization. We are planning to develop a series of short persuasive messages explaining the system, its background, and the need for it.

Second, the user research described in this article was, in line with the state of development of the Privacy Rating, limited to specific aspects of the label after artificial exposure. In the usability test, we focused predominantly on the perceived understandability of the various elements of the Privacy Rating. It would be interesting in follow-up research to focus more on participants' interpretations and actual use of the label as a whole. The research into the effects of the Privacy Rating was limited to the explicit question whether the participants would trust the web shop enough to do business with it. Follow-up research in a more natural setting would less exclusively and explicitly focus on the trust question and ask how the privacy label might, for instance, affect the image or reputation of the online service provider. The question whether a positive Privacy Rating can be good for business would be very relevant, as it could convince online service providers to embrace transparency regarding privacy and include the Privacy Rating in their communication.

Third, experimental research is needed to further investigate the two tradeoffs that we mentioned: between conciseness/simplicity and informativeness (which balance is most effective for which user groups) and between annoying intrusiveness and sheer invisibility (how can we make a privacy label optimally visible without annoying users). Finally, it would be interesting to extend research in laboratory settings with real-life research into the users' appreciation of and behavior toward the Privacy Rating.

Conclusion

We propose Privacy Rating, which addresses the inherent functional complexity of privacy communication by visually synthesizing information across multiple layers of increasing detail. It thereby increases awareness, provides less engaged users with shortcuts, and supports privacy aware users in making informed decisions. Usability testing showed that the label was perceived as useful and usable. It also had a significant effect on trust in the online service. All participants indicated they would appreciate such a label becoming an established standard. More generally, we learned that privacy visualizations should use familiar design elements and ideally be supported by a trustworthy organization.

In our future work, we will concentrate on refining the Privacy Rating and making it market-ready. We are therefore looking for prospective partners and organizations interested in future collaboration to bring the Privacy Rating to the market. A privacy visualization that satisfies these requirements can empower users by significantly improving privacy awareness and helping achieve truly informed consent.

Acknowledgment

This work was supported by the Dutch Research Council (NWO; grant number: 628.001.011) in collaboration with the Netherlands Organisation for Applied Scientific Research (TNO), the Dutch Research and Documentation Centre (WODC), and Centric.

References

[1] T. Berners-Lee, J. Handler, and O. Lassila, "The semantic web," *Sci. Amer.*, vol. 284, no. 5, pp. 34–43, 2001.
[2] G. R. Milne, A. J. Rohm, and S. Bahl, "Consumers' protection of online privacy and identity," *J. Consum. Affairs*, vol. 38, no. 2, pp. 217–232, 2004.
[3] K. Huckvale, J. T. Prieto, M. Tilney, P.-J. Benghozi, and J. Car, "Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment," *BMC Med.,* vol. 13, no. 1, 2015, Art. no. 214.
[4] S. C. Matz, R. E. Appel, and M. Kosinski, "Privacy in the age of psychological targeting," *Curr. Opin. Psychol.*, vol. 31, pp. 116–121, 2020.
[5] M. Alsaleh, N. Alomar and A. Alarifi, "Smartphone users: Understanding how security mechanisms are perceived and new persuasive methods," *PLoS ONE*, vol. 12, no. 3, 2017, Art. no. e0173284.
[6] J. A. Mourey and A. E. Waldman, "Past the privacy paradox: The importance of privacy changes as a function of control and complexity," *J. Assoc. Consum. Res.*, vol. 5, no. 2, pp. 162–180, 2020.

[7] M. Rudolph, D. Feth, and S. Polst, "Why users ignore privacy policies. A survey and intention model for explaining user privacy behavior," in *Human-Computer Interaction: Theories, Methods, and Human Issues*. M. Kurosu, ed. Cham, Switzerland: Springer, 2018, pp. 587–598.
[8] R. W. Proctor, M. A. Ali, and K.-P. L. Vu, "Examining usability of web privacy policies," *Int. J. Hum. Comput. Interact.*, vol. 24, no. 3, pp. 307–328, 2008.
[9] J. Prichard and K. Mentzer, "An analysis of app privacy statements," *Issues Inf. Syst.*, vol. 18, no. 4, pp. 179–188, 2017.
[10] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2004, pp. 179–188.
[11] A. Sunyaev et al., "Availability and quality of mobile health app privacy policies," *J. Amer. Med. Inf. Assoc.*, vol. 22, pp. e28–e33, 2015.
[12] Z. Papacharissi and J. Fernback, "Online privacy and consumer protection: An analysis of portal privacy statements," *J. Broadcast. Electron. Media*, vol. 49, no. 3, pp. 259–281, 2005.
[13] A. Acquisti et al., "Nudges for privacy and security: Understanding and assisting users' choices online," *ACM Comput. Survey*, vol. 50, no. 3, 2017, Art. no. 44.
[14] J. Gu et al., "Privacy concerns for mobile app download: An elaboration likelihood model perspective," *Decis. Support Syst.*, vol. 94, pp. 19–28, 2017.
[15] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford, CA, USA: Stanford Univ. Press, 1957.
[16] S. Chai, S. Bagchi-Sen, C. Morrell, H. R. Rao, and S. J. Upadhyaya., "Internet and online information privacy: An exploratory study of preteens and early teens," *IEEE Trans. Profess. Commun.*, vol. 52, no. 2, pp. 167–182, Jun. 2009.
[17] S. Young, "Zoombombing your toddler: User experience and the communication of zoom's privacy crisis," *J. Bus. Tech. Commun.*, vol. 35, no. 1, pp. 147–153, 2021.
[18] P. G. Kelley et al., "A 'nutrition label' for privacy," in *Proc. 5th Symp. Usable Privacy Secur.*, 2009, Art. no. 4.
[19] P. G. Kelley et al., "Standardizing privacy notices: An online study of the nutrion label approach," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 1573–1582.
[20] L.-E. Holtz, K. Nocun and M. Hansen, "Towards displaying privacy information with icons," in *Privacy and Identity Management for Life*. S. Fischer-Hübner et al., eds. Heidelberg, Germany: Springer, 2011, pp. 338–348.
[21] L.-E. Holtz, H. Zwingelberg, and M. Hansen, "Privacy policy icons," in *Privacy and Identity Management for Life*. J. Camenisch et al., eds. Heidelberg, Germany: Springer, 2011, pp. 279–285.
[22] B. van den Berg and S. van der Hof, "What happens to my data? A novel approach to informing users of data processing practices," *First Monday*, vol. 17, no. 7, 2012.
[23] J. Petterson, "A brief evaluation of icons in the first reading of the European Parliament on COM (2012) 0011," in *Privacy and Identity Management For the Future Internet in the Age of Globalisation*. J. Camenisch et al., eds. Heidelberg, Germany: Springer, 2012, pp. 125–135.
[24] L. Edwards and W. Abel. (2014-2015). The use of privacy icons and standard contract terms for generating consumer trust and confidence in digital services. CREATe working paper. Centre for Copyright and New Business Models in the Creative Economy (CREATe), 2014, Glasgow, UK. [Online]. Available: https://zenodo.org/record/12506/files/CREATe-Working-Paper-2014-15.pdf
[25] S. Esayas et al., "Is a picture worth a thousand terms? Visualising contract terms and data protection requirements for cloud computing users," in *Current Trends in Web Engineering*. S. Casteleyn et al., eds. Cham, Switzerland: Springer, 2016, pp. 39–56.
[26] A. Rossi and M. Palmirani, "A visualization approach for adaptive consent in the European data protection framework," in *Proc. Int. Conf. E-Democracy Open Govern.*, Krems, Austria, 2017, pp. 159–170.
[27] A. Rossi and M. Palmirani, "DAPIS: An ontology-based data protection icon set," in *Knowledge of the Law in the Big Data Age*. G. Peruginelli and S. Faro, eds. Amsterdam, the Netherlands, IOS Press, 2019, pp. 181–195.
[28] M. Van Kleek et al., "Better the devil you know: Exposing the data sharing practices of smartphone apps," *Proc. CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 5208–5220.
[29] K. Renaud and L. A. Shepherd, "How to make privacy policies both GDPR-compliant and usable," in *Proc. Int. Conf. Cyber Situational Awareness, Data Analytics Assessment*, 2018.
[30] W. B. Tesfay et al., "PrivacyGuide: Towards an implementation of the EU GDPR on internet privacy policy evaluation," in *Proc. 4th ACM Int. Workshop Secur. Privacy Analytics*, 2018, pp. 15–21.
[31] G. Fox et al., "Communicating compliance: Developing a GDPR privacy label," in *Proc. 24th Am. Conf. Inf. Syst.*, Atlanta, GA, USA: Association for Information Systems, 2018.
[32] Z. Efroni et al., "Privacy icons: A risk-based approach to visualisation of data processing," *Eur. Data Prot. Law Rev.*, vol. 5, no. 3, pp. 352–366, 2019.
[33] A. Soumelidou and A. Tsohou, "Effects of privacy policy visualization on users' information privacy awareness level: The case of Instagram," *Inf. Tech. People*, vol. 33, no. 2, pp. 502–534, 2019.
[34] S. de Jong and D. Spagnuelo, "Iconified representations of privacy policies: A GDPR perspective," in *Trends and Innovations in Information Systems and Technologies*. vol. 2, A. Rocha et al., eds. Cham, Switzerland: Springer, 2020, pp. 796–806.
[35] The European Parliament and the Council of European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 On the Protection of Natural Persons With Regard to the Processing of Personal Data and On the Free Movement of Such Data, and Repealing Directive 95/46/EC*, vols. Legislation OJ L vol. 119, 2.5, Luxembourg, Europe: Publ. Office Eur. Union, 2016.
[36] S. B. Barnes, "A privacy paradox. Social networking in the United States," *First Monday*, vol. 11, no. 9, 2006, doi: 10.5210/fm.v11i9.1394.

[37] S. Barth and M. D. T. de Jong, "The privacy paradox. Investigating discrepancies between expressed privacy concerns and actual online behavior. A systematic literature review," *Telematics Inf.*, vol. 34, no. 7, pp. 1038–1058, 2017.
[38] M. Langheinrich, "Privacy by design. Principles of privacy-aware ubiquitous systems," in *Ubicomp 2001: Ubiquitous Computing*. G. D. Abowd, B. Brumitt, and S. Shafer, eds. Berlin, Germany: Springer, 2001, pp. 273–291.
[39] A. Cavoukian. (2011 Jan.). Privacy by design. The 7 foundational principles. Tech. rep. (revised version), Information and Privacy Commissioner of Ontario. [Online]. Available: http://dataprotection.industries/wp-content/uploads/2017/10/privacy-by-design.pdf
[40] ISO/IEC 29100:2011, *Information Technology—Security Techniques—Privacy Framework*. Geneva, Switzerland: International Organization for Standardization and International Electrotechnical Commission. [Online]. Available: https://www.iso.org/standard/45123.html
[41] B.-J. Koops and R. Leenes, "Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law," *Int. Rev Law, Comput. Tech.*, vol. 28, no. 2, pp. 159–171, 2014.
[42] D. Klitou, "A solution, but not a panacea for defending privacy: The challenges, criticism and limitations of privacy by design," in *Annual Privacy Forum*, B. Preneel and D. Ikonomou, eds. Berlin, Germany: Springer, 2012, pp. 86–110.
[43] F. Bu et al., "Privacy by design implementation: Information system engineers' perspective," *Int. J. Inf. Manage.*, vol. 53, Art. no. 102124, 2020.
[44] A. A. Gerunov, "Attitudes towards privacy by design in e-government: Views from the trenches," *J. Soc. Admin. Sci.*, vol. 7, no. 1, pp. 1–17, 2020.
[45] A. Cavoukian, "Understanding how to implement privacy by design, one step at a time," *IEEE Consum. Electron. Mag.*, vol. 9, no. 2, pp. 78–82, Mar. 2020.
[46] Forbrukerrådet, *Deceived By Design, How Tech Companies Use Dark Patterns to Discourage Us from Exercising Our Rights to Privacy*. Oslo, Norway: Forbrukerrådet [Consumer Council], 2018.
[47] S. Barth et al., "Toward an understanding of online privacy perceptions: Using the Q-sort method to identify different user perspectives," 2021.
[48] S. Barth et al., "Putting the privacy paradox to the test. Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources," *Telematics Inform.*, vol. 41, no. 1, pp. 55–69, 2019.
[49] S. Barth et al., "Lost in privacy? Online privacy from a cybersecurity expert perspective," 2021.
[50] O. Ben-Shahar and A. Chilton, "Simplification of privacy disclosures: An experimental test," *J. Legal Studies*, vol. 45, no. S2, pp. S41–S67, 2016.
[51] R. LaRose and N. J. Rifon, "Promoting i-safety: Effects of privacy warnings and privacy seals on risk assessment and online privacy behavior," *J. Consum. Affairs*, vol. 41, no. 1, pp. 127–149, 2007.
[52] L. Lentz and H. Pander Maat, "Functional analysis for document design," *Tech. Commun.*, vol. 51, no. 3, pp. 387–398, 2004.
[53] M. D. T. de Jong and Y. Wu, "Functional complexity and web site design: Evaluating the online presence of UNESCO world heritage sites," *J. Bus. Tech. Commun.*, vol. 32, no. 3, pp. 347–372, 2018.
[54] A. Deuker, "Addressing the privacy paradox by expanded privacy awareness. The example of context-aware services," in *Privacy and Identity Management For Life*. M. Bezzi et al., eds. Berlin, Germany: Springer, 2009, pp. 275–283.
[55] S. Pötzsch, "Privacy awareness: A means to solve the privacy paradox?," in *The Future of Identity*. V. Matyáš et al., eds. Berlin, Germany: Springer, 2008, pp. 226–236.
[56] X. Sheng et al., "Sight unseen: The role of online security indicators in visual attention to online privacy information," *J. Bus. Res.*, vol. 111, pp. 218–240, 2020.
[57] S. Barth et al., "Understanding online privacy: A systematic review of privacy visualizations and privacy by design guidelines," 2021.
[58] M. Mehldau. Iconset for data-privacy declarations. v0.1. 2007. [Online]. Available: https://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf
[59] J. Gomez, T. Pinnick, and A. Soltani. (2009). KnowPrivacy. UC Berkeley, School Inf., Berkeley, CA, USA. [Online]. Available: http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf
[60] B. Moskowitz and A. Raskin. (2011). Privacy icons. Mozilla Wiki. [Online]. Available: https://wiki.mozilla.org/Privacy_Icons
[61] C. Graf et al., eds., (2011). PrimeLife—Privacy and identity management in Europe for life: Final HCI Res. Rep., 2011. [Online]. Available: http://primelife.ercim.eu/images/stories/deliverables/d4.1.5-final_hci_research_report-public.pdf
[62] T. Pinnick, *Privacy Short Notice Design*. San Francisco, CA, USA: TrustArc, 2011. [Online]. Available: https://trustarc.com/blog/2011/02/17/privacy-short-notice-design/
[63] G. Franke, T. Clever, W. van Dijk, J. Raider, and R. de Jonge, *Privacy Label*. Blog series part I-IV. Sensor lab, 2019. [Online]. Available: https://medium.com/sensor-lab/the-privacy-illusion-994ed98ec3ab
[64] *Upgrade Your Privacy Statement: Privacy Label Helps You Communicate How You Use Data*. Privacy Label, 2020. [Online]. Available: https://www.privacylabel.org/learn/how-does-it-work
[65] M. Hansen, "Putting privacy pictograms into practice. A European perspective," in *Informatik 2009–Im Focus das Leben*. S. Fischer et al., ed. Bonn, Germany: Gesellschaft für Informatik, 2009, pp. 1703–1716.
[66] H. Nissenbaum, "A contextual approach to privacy online," *Daedalus*, vol. 140, no. 4, pp. 32–48, 2011.

[67] H. Nissenbaum, "Contextual integrity up and down the data food chain," *Theoritical Inquiries Law*, vol. 20, no. 1, pp. 221–256, 2019.
[68] European Parliament and Council of European Union. *Regulation (EU) 2017/1369 of the European Parliament and of the Council of 4 July 2017 Setting a Framework For Energy Labelling and Repealing Directive 2010/30/EU. Legislation OJ L 198, 28.7.2017*. Luxembourg, Luxembourg: Publ. Office Eur. Union, 2017. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1369&rid=4
[69] European Commission, *Energy Label Templates*. Directorate-General for Energy, Brussels, Belgium, 2020. [Online]. Available: https://ec.europa.eu/energy/topics/energy-efficiency/energy-label-and-ecodesign/energy-label-templates_en?redir=1
[70] R. Balebako et al., "The impact of timing on the salience of smartphone app privacy notices," in *Proc. 5th Annu. ACM CCS Workshop Security Privacy Smartphones Mobile Devices*, 2015, pp. 63–74.

Susanne Barth received the Ph.D. degree in Communication Science from the University of Twente, Enschede, the Netherlands, in 2021. Her Ph.D. project centered on privacy and security requirements for mobile applications. She is a Postdoctoral Researcher within the Communication Science Group, BMS faculty, and the Services and CyberSecurity group, EEMCS faculty at the University of Twente. Her research expertise is strongly rooted within communication science and psychology, primarily with a focus on the human factor in technology and societal challenges intertwined with new media and technologies these days.

Dan Ionita received the bachelor's degree in Electrical Engineering from the Polytechnic University of Bucharest, in 2011 and the M.Sc. degree in Information Systems Engineering and the Ph.D. degree in Information Security Risk Assessment from the University of Twente, Enschede, the Netherlands, in 2013 and 2018, respectively. He is a Postdoctoral Researcher within the Services and Cyber-Security Department of the University of Twente. He is currently researching, teaching, and providing consultancy in the field of privacy and information security.

Menno D. T. de Jong received the Ph.D. degree in Communication Science from the University of Twente, Enschede, the Netherlands, in 1998. He is currently a Full Professor of communication science with the University of Twente. Between 2009 and 2015, he was the Editor-in-Chief of *Technical Communication*, the flagship journal of the Society for Technical Communication. He has authored or coauthored in a broad range of academic journals. His research interests include the fields of technical and organizational communication. Prof. de Jong was the recipient of various awards for his research, including two Rudolph J. Joenk, Jr. awards from the IEEE Professional Communication Society for Best Paper in the TRANSACTIONS, the Ken Rainey Award for Excellence in Research from the STC, and the Alfred N.Goldsmith Award for Outstanding Achievement in Engineering Communication from the IEEE Professional Communication Society.

Pieter H. Hartel received the Ph.D. degree in Computer Science from the University of Amsterdam, Amsterdam, the Netherlands, in 1989. He is Professor Emeritus of cyber security with the Delft University of Technology, Delft, the Netherlands, and the University of Twente, Enschede, the Netherlands, and an Adjunct Faculty with the Singapore University of Technology and Design. His research interests include cybersecurity, cybercrime, and block-chain technology.

Marianne Junger received the Ph.D. degree in law from the Free University of Amsterdam, Amsterdam, the Netherlands, in 1990. She is the Emeritus Professor of Cyber Security and Business Continuity with the University of Twente, Enschede, the Netherlands. Her research investigates the human factors of fraud and cybercrime. More specifically, she investigates victimization, disclosure, and privacy issues. She founded the *Crime Science* journal together with Pieter Hartel and was an Associate Editor for 6 years. Her research was sponsored by, among others, the Dutch Police, NWO, ZonMw (for health research), and the European Union.

Research Article

User Perception of Data Breaches

—Zahra Hassanzadeh, Robert Biddle, and Sky Marsen

***Abstract*—Background:** *Data breaches happen when an unauthorized party gains access to personally identifiable information. They are becoming more common and impactful, raising serious concerns for individuals as well as companies.* **Literature review:** *Although there is considerable literature on users' mental models in security and privacy, there has been limited study of mental models related to data breaches.* **Research questions:** *1. How do users understand data breaches? 2. What are their perceptions of the causes, responsibilities, and consequences, as well as possible prevention and appropriate follow up?* **Methodology:** *We explored end-user understanding of internet data breaches by conducting a study with 35 participants. They were asked to draw their understanding of data breaches and answer some open-ended and closed-ended questions afterwards.* **Results/discussion:** *Although their drawings varied in detail and complexity, we identified four patterns in the participants' drawings: they illustrated abstractions of attacks to gain administrator access, end-user access, backdoor access, or access using database server vulnerabilities. We found that participants had a basic model of how an internet data breach happens, but with significant uncertainties regarding system vulnerabilities, causes, consequences, prevention methods, and follow-up steps after a breach.* **Conclusions:** *In all, end-user mental models of internet data breaches are basic and show gaps that emphasize the need for improved communication to increase users' awareness and help them hold companies accountable.*

Index Terms*—Cybersecurity, internet data breaches, mental models.*

Data breach incidents are becoming more common and have aroused broad concern. According to the *Risk Based Security* report, a global database of public data breaches, over 8 billion records were exposed in the first quarter of 2020 [1]. A data breach is an incident in which cybercriminals attack a system, infiltrate a data source, and extract personally identifiable information that can be helpful to distinguish an individual's identity [2]. All computer attacks do not necessarily lead to a data breach. However, attackers may probe to find system vulnerabilities and compromises that allow them to get access to companies' servers and expose their data. Attackers can also use targeted malware attacks to gain access to a target system.

Intrusion detection systems (IDS) are deployed to defend companies from the breaches and attacks that arise on a daily basis. IDS monitor a network or system for malicious activity. Generally, they will notify administrators of a possible intrusion or collect the malicious activity centrally using security information and an event management system. However, IDS are not completely effective.

Researchers in cybersecurity emphasize the role of security mechanisms such as firewalls, authentication mechanisms, virtual private networks, and security technology such as IDS, which are complementary to security mechanisms [3]. Despite increased awareness and improvement of security mechanisms, companies still practice insufficient cybersecurity measures, which cause breaches such as the breaches of Equifax in 2017 [4], Marriott International in 2014–2018, LinkedIn in 2012 and 2016, Yahoo in 2013–2014, and many more [5].

Despite the increase in cyberbreaches, there has been surprisingly little engagement from the public, and many people continue using services after breaches. Some suggest that notable data breaches do not change customers' behavior, attributing it "breach fatigue" [6]. Users' understanding of how things work is important in making security decisions because it pinpoints the mental models that shape the way that they see the breach [7]–[9]. Therefore, finding the gaps in the users' mental models would indicate how communication can be

Manuscript received December 1, 2020; revised March 3, 2021; accepted April 8, 2021. Date of publication October 19, 2021; date of current version November 24, 2021. *(Corresponding author: Zahra Hassanzadeh.)*
Zahra Hassanzadeh is with Public Services and Procurement Canada, Ottawa, ON K1A 0X1, Canada (email: zahrahassanzadeh@cmail.carleton.ca).
Robert Biddle is with the School of Computer Science, Carleton University, Ottawa, ON K1S 5B6, Canada (email: robert.biddle@carleton.ca).
Sky Marsen is with Flinders University College of Humanities, Arts and Social Sciences, Adelaide 5001, Australia (email: skymarsen@gmail.com).

IEEE 10.1109/TPC.2021.3110545

Practitioner Takeaway

- Non-specialist users demonstrate inadequate understanding of how and why data breaches happen, and consequently are not able to attribute responsibility appropriately.
- The concept of software security vulnerability, which is vital in understanding data breaches, is actually misunderstood by non-specialist users, who often confuse it with its psychological use.
- As a consequence of the above, non-specialist users have difficulty in managing a situation where their information has been breached and may be unaware of suitable actions.

designed to help users reach reasonable security expectations for companies to keep their data safe.

Considerable literature in computer security and privacy addresses mental models; however, research on mental models related to data breaches is rare. The goal of our research was to address this gap by exploring user misunderstanding of internet data breaches that endangers the user's personally identifiable information. Our objective was to investigate some problematic beliefs about data breaches in users' mental models to trace patterns that could be used to inform the design of relevant communication materials produced by organizations and regulatory authorities. As well as contributing generally to research on the mental models of users related to data breaches, this project aimed specifically to pinpoint certain user misconceptions about data breaches that risk and crisis communicators in organizational, regulatory, and instructional contexts should consider.

To this end, we conducted a study in which we asked users to describe and illustrate their understanding of how a data breach occurs. We collected data about participants' perceptions of internet data breaches; their perceptions of the causes, motivations, and consequences of those breaches; as well as their attitudes toward preventing a data breach. This study complements our earlier research on image-repair strategies following a data breach [10], which focused on the public communication of organizations and media responses after a data breach. That study explored the ways that organizations attempt to influence and direct users' and more generally the public's perception of the breach so as to reduce organizational accountability.

Literature Review

Mental models are explanatory representations that humans develop when interacting with environments, with others, and with technology. Mental models are employed by discourse analysts, psychologists, and sociologists to identify people's problem-solving techniques based on the way they perceive the problem and the context in which it arises [7], [11]. A mental model approach has been employed in crisis communication research to study the public's reactions to crisis events and open the way for the design of more effective communication practices. For example, Morgan et al. [12] explain the importance of communicating risk based on the public's existing beliefs and assumptions and not on what "experts" think that they should know.

Regarding data breaches, most users' decisions are shaped by their mental models [13], which act like filters affecting the way they see and interpret the world [7]–[9]. Because mental models show established ways of interaction with a system, they can be useful in designing protocols for communicating security-related concepts and processes to nonspecialist users.

Jean Camp [14] presented five possible mental models for security failures by using metaphors: physical security, medical infection, criminal behavior, economic failure, and warfare. Each of these models implies a different solution, and therefore, when communicating with users, computer security experts should consider which model will meet user expectations.

- Physical security models are somewhat understood; for instance, the metaphor of a wall might help to understand a network security practice. This metaphor encourages users to secure their computers.
- The public health model reminds users to protect themselves and draws on the need for shared responsibility for community health.
- In the criminal behavior model, malicious behavior is a crime and can result in victims suffering a significant loss. So law enforcement is a logical response, and increased surveillance is a prominent part of the solution to the computer crime problem.

- The warfare model evokes the existence of a determined enemy and the critical need for response and reminds the users that individual actions are crucial for collective security.
- Computer security failures can also result in economic failure as they are costly to an organization. So this model would justify convincing users that they have valuable assets and that thieves might target them.

Researchers use different methods to identify mental models, including problem solving, verbal reports, drawing, categorization, and conceptual pattern representation. In the problem-solving method, people use their mental models to understand a problem and make inferences and predictions [15], so those with different mental models will understand a problem differently and will come up with different solutions. A verbal report is a direct method of eliciting mental models and can encompass interviews, explanations, or think-aloud protocols [16]. In think-aloud protocols, participants are asked to verbalize their thoughts while performing a task. Think-aloud is useful for analyzing mental models because it provides direct information about the participants' thinking process. Drawing is frequently used in conjunction with verbal reports [17] in user-centered research, and has also been found to assist in elicitation of mental models [18]. The categorization method shows how mental models are developed and categorized, and has usually been used to show the similarities and differences between experts and novices [19]. Finally, conceptual patterns represent the concepts and their relationships [20], [21].

Because mental models are often represented pictorially and using verbal reports has some limitations [17], drawings have been used as complementary methods in several mental model research studies. For example, Raja et al. [22] conducted a study of participants' mental models of the Vista Firewall using a diagramming task. They gave participants a picture of a computer, a firewall, and the internet cloud and asked them to show how the Vista Firewall works by drawing arrows. Subsequently, they designed a new interface with contextual information to improve participants' mental models of the firewall.

Wash [8] used multiple rounds of interviews to understand users' mental models of security threats. He identified eight folk models of malware and attackers. Users employed these models to decide which security software to use and which advice to follow. Wu and Zappala [23] conducted a series of semistructured phone interviews using interview techniques and diagramming exercises to capture users' perception of the encryption process. They found four mental models that are different in detail and complexity. They also found that users do not fully understand the decisive role of encryption.

Clearly, understanding user mental models can lead to better communication regarding privacy and security risks [24]. Evidence suggests that deficient mental models of data breaches influence behavior, and people tend to underestimate the consequences and risks of a data breach. Ideally, users need to be able to answer these questions:

- What is a data breach?
- How do data breaches happen?
- What are the consequences of a data breach?
- How can individuals and companies prevent data breaches?
- What are a company's responsibilities regarding securing personal data?
- What can users do after a data breach?

Research Questions

The following research questions guided our study.

RQ1. How do users understand data breaches?
RQ2. What are their perceptions of the causes, responsibilities, and consequences, as well as possible prevention and appropriate follow up?

The objective of these questions is to form a framework that will inform communication practices and interventions by organizations and other relevant parties aimed at dispelling misconceptions and empowering users when they make security decisions.

Methodology

The data presented in this article are derived from 35 one-on-one, in-person sessions with participants about their mental models of internet data breaches. Each session lasted 40 to 90 minutes. Participants were compensated with a small monetary payment. Our study was reviewed and cleared by a university Research Ethics Board.

We collected data from a drawing task, as well as open-ended and closed-ended questions administered through LimeSurvey (an open source survey system, hosted on a secure server). We audio-recorded participants' voices while they were verbally explaining their understanding of internet data breaches. These recordings were

TABLE I
PARTICIPANTS' DEMOGRAPHICS (TOTAL= 35)

Age	%	Education	%	Field of study	%
18-30	74%	Less than High school	0%	Formal science (Computer Science, Logic, Math)	17%
31-40	20%	High school degree or equivalent	18%	Natural science (Biology, Physics, Chemistry…)	11%
41-50	3%	College/Bachelor's degree	51%	Social science	29%
51-60	3%	Trade or technical degree	3%	Engineering	14%
Over 60	0%	Graduate degree	28%	Arts	17%
				Law	6%
				Other	6%

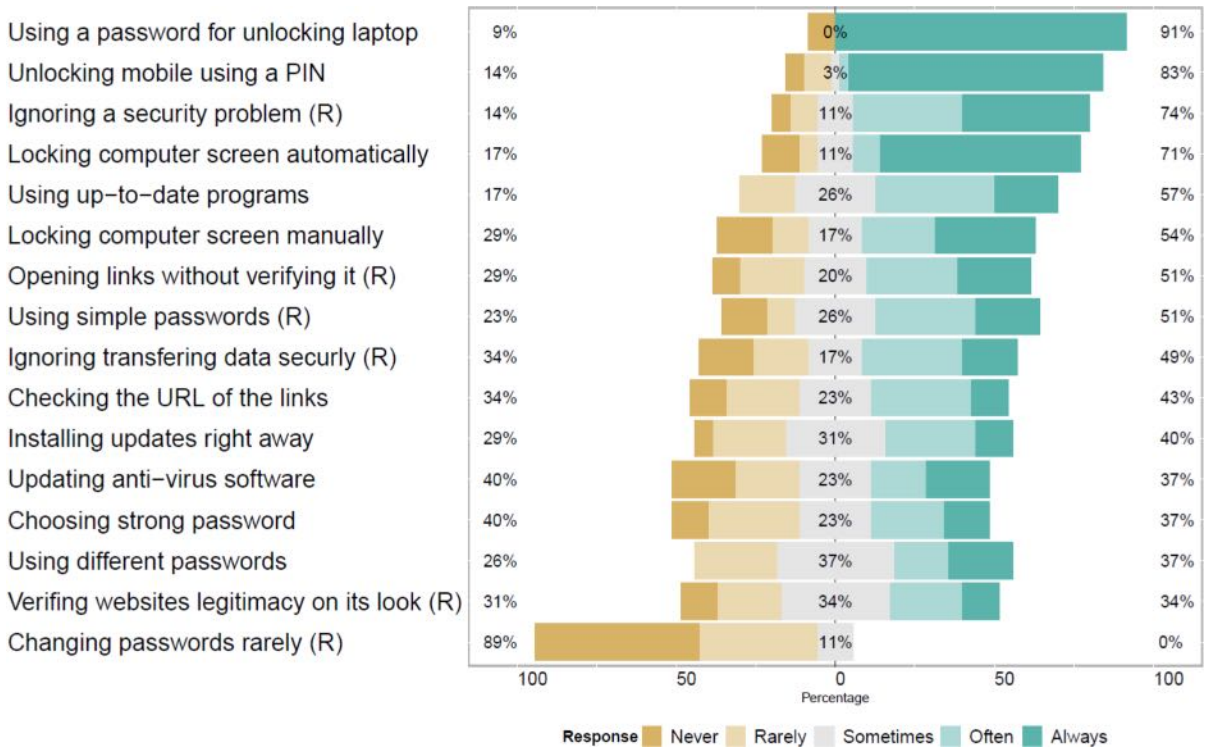


Fig. 1. SeBIS results. Statements marked (R) were reverse scored.

transcribed using Otter.ai (https://otter.ai), and the transcriptions were reviewed during our analysis.

Participants Participants were recruited using recruitment posters, and a Facebook page for user study recruitment; known contacts were also emailed. We recruited a total of 35 participants: 18 female and 17 male. They ranged from 18 to 54 years (see Table I). When asked specifically if they have ever experienced any data security problem, 11 said yes, 18 said no, and the remaining six said that they were uncertain. When asked whether they had ever experienced identity theft, 27 said no, five said yes, and the remaining three said that they were uncertain. Of the 35 participants, 34 agreed to be audio-recorded, and one participant did not, so we took notes during that session.

Sessions The sessions took place in a research lab or an offsite location. Participants read and signed a consent form after we explained the tasks. They first completed a basic demographic, and a Security Behavior Intentions Scale (SeBIS) questionnaire [25]. Then, participants were asked

to draw their understanding of how a data breach happens in response to the following prompt.

Using a paper and pencil, please draw your understanding of how a data breach happens. In other words, try to draw what is happening when it is said that an attacker has hacked a company's database servers and personal information about you *and many other people has been breached.* (Elements you may need for this drawing: Admin, Data, Customer, Hacker, Access to data, Vulnerability).

We made a change after the first 15 participants' sessions to emphasize the breadth of the attack (see italics) but did not observe any changes in participant responses.

After the drawing task, participants answered open-ended questions and Likert-type questions on a 7-point scale regarding causes, consequences, motivations, and prevention methods of a data breach. After reviewing the result of the Likert-type questions of the early round of study (15 participants), we speculated that participants answered some questions simply agreeing with anything related to security because we were interested in that topic. So in the second round of study, we added more plausible but inappropriate items as well as clearly wrong answers to check and ensure the validity of their responses. The participants' responses changed in the second round, so we reported only these data (20 participants) because they are more reliable.

Data Analysis We addressed participant responses in two sections: qualitative data, and quantitative data stemming from Likert-scale responses. We analyzed the results of the SeBIS questionnaires by aggregating the score of positive questions minus the score of negative ones. Each question was scored 1–5, where 1 stands for "Never," 3 for "Sometimes," and 5 for "Always." The other closed answers were analyzed by computing the median for the central tendency of our data, as well as analyzing and reviewing graphical representations.

The high-level codes for open questions were associated with our research questions and were focused on the following.

- What is considered a data breach?
- What are the causes of an internet data breach?
- What reasons are given for an internet data breach?
- How do you know whether a user has been affected in an internet data breach?
- What are the consequences and prevention methods?
- Who is responsible for data protection?

These themes are top-level categories, and we created several subcategories based on participants' responses.

We then conducted a thematic analysis [26] of participants' think-aloud comments during their drawing task, discussion of their drawings, and responses to open-ended questions. We started by examining the diagrams closely to understand the data as a group. We were looking for key elements, trends, themes, or ideas in the images. We generated a set of codes by comparing the similarities and differences between the participants' drawings. The codes are representative of the identified themes and are linked to drawings as summary markers [27], [28]. We built an initial summary of these codes and identified patterns in the ways participants talked about data breaches and their word choices, looking for cases where they had different perceptions. We used the transcriptions as a complementary source for the drawing task to better understand the drawings.

Results

SeBIS Question Results Fig. 1 shows a summary of the results of the SeBIS questionnaire, illustrating the computer security behaviors of participants. Notably, most reported risky behavior regarding password generation: they only sometimes used different passwords for different accounts (37%), they only sometimes used a password that exceeded the site's minimum requirements (37%), and they often did not include special characters in their passwords when not required (51%).

When asked about proactive awareness, they reported risky behavior like sometimes opening a link without first looking to see where it went. They sometimes submitted information to websites without first verifying that it would be sent securely, and if they discovered a security problem, they often (74%) continued what they were doing because they assumed someone else would fix the problem. They reported risky behavior regarding updating software and antivirus programs (57%).

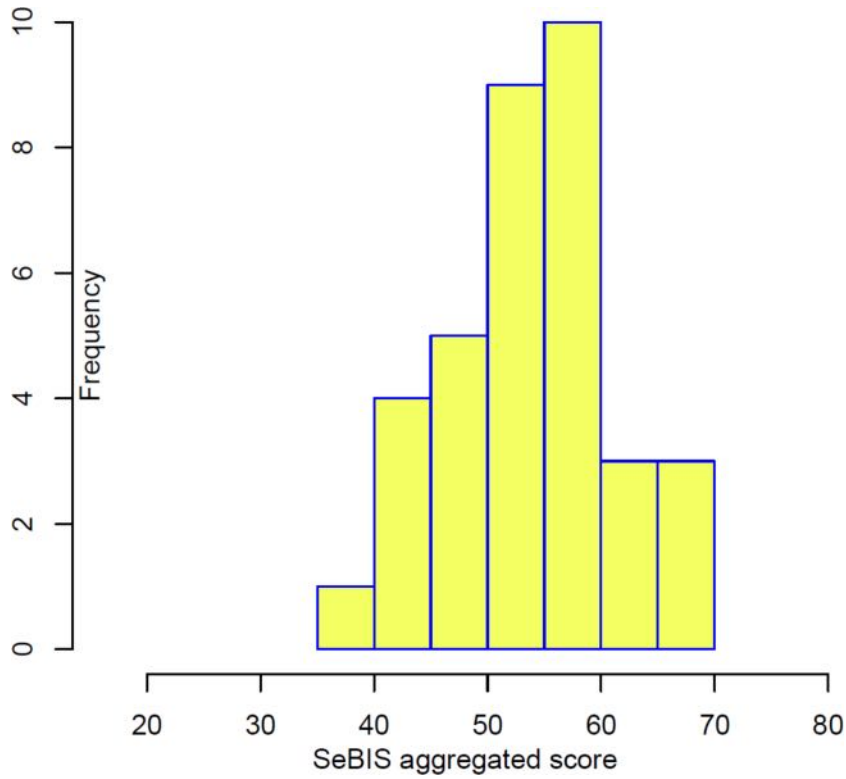


Fig. 2. SeBIS histogram; maximum score is 80 and minimum is 16.

The histogram in Fig. 2 shows their different levels of aggregate security behavior. The possible range of participant scores was between 0 and 80, and the plot illustrates a reasonable distribution, showing that the participants were neither security "advocates" nor security "slackers."

Questionnaire Results At the end of each session, participants were asked to answer another questionnaire consisting of some open-ended questions and a series of 7-point Likert-scale questions that were grouped to represent the following categories.

1. Definition of a data breach and responsibility for data protection
2. Causes of data breaches
3. Consequences of a data breach
4. Motivation for data breaches
5. Prevention methods
6. Actions after a breach

1. *Definition of a Data Breach and Responsibility for Data Protection:* In response to being asked what a data breach is, all participants had a general understanding of how an unauthorized party gets access to the secure or private/confidential information. For example, P11 defined a data breach as "… when information is kept privately on the internet and is then accessed by someone who is not supposed to have access to the information."

In response to being asked whether they had ever heard of an internet data breach, 25 participants said yes, three said no, and seven said they were not sure. In response to how they had heard of data breach incidents, they mentioned news media, breach notifications, and word of mouth. When asked how they would know whether they had been affected in an internet data breach, participants showed a reasonable understanding of how users become aware of breaches (see Table II).

When asked who is responsible for data protection, 28 participants considered the users responsible for protecting data. When asked the same question in a closed-ended format, 60% of participants believed that government and customers of a company are responsible for data protection. For this and some other questions, early results suggested guessing, so we offered several options with seemingly plausible but inappropriate answers. As a result, more than 40% of participants said that they either believed or were uncertain whether the American National Standard Institute, the Group of 7 major advanced nations (G7), General Agreement on Tariffs and Trade (GAAT), or Organization for Economic Cooperation

TABLE II
HOW WOULD YOU KNOW IF YOU HAVE BEEN AFFECTED IN AN INTERNET DATA BREACH?

Code	Total	Sample of Data Extracted
Company's notifications	13	"I would hopefully be contacted by the company who lost my information and be told what was taken."—P2
Alerts to my accounts	9	"When you are notified about your activities (transactions, logging in, etc.) on internet through SMS or email notification."—P19
Checking my bank transactions	9	"This can be seen in many ways, such as a sudden change in your bank statement that looks bizarre and unusual purchases have taken place." —P15
Never know	7	"You probably wouldn't know if the attackers did a good job."—P5
News	8	"I would know I was affected in an internet data breach because I might hear about it through the media."—P2
Removed info	3	"My information on the phone gets removed."—P8
Blackmailing victim	3	"Maybe noticing data being used in a way you don't remember doing? Or they'll tell you, if they're hoping the breach will make you do something for them."—P5

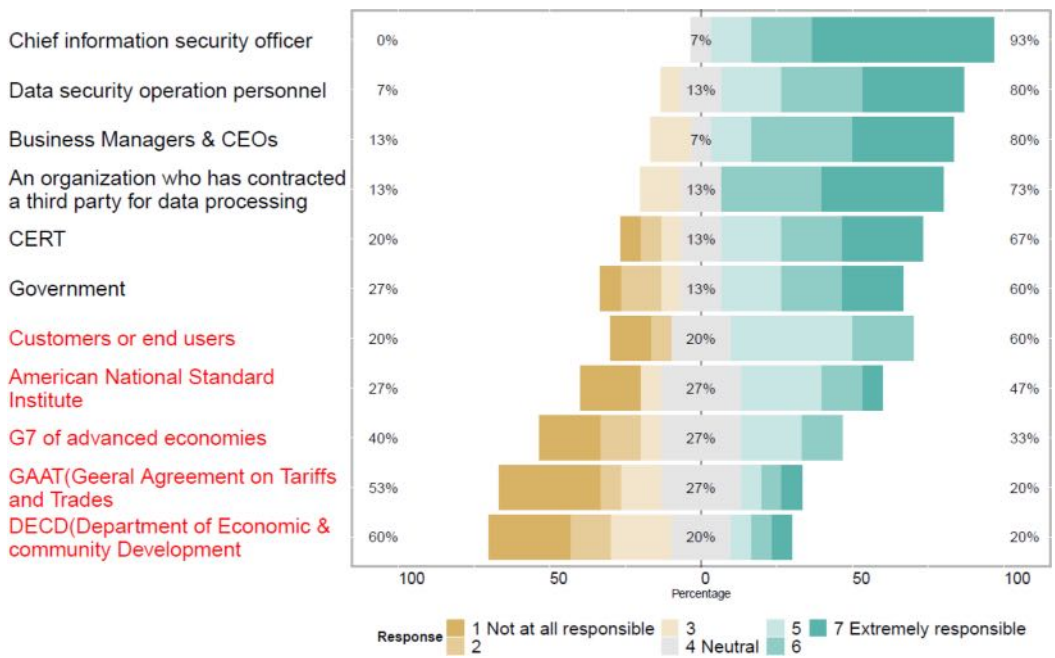


Fig. 3. Accountability of people for data protection. Labels in red are inappropriate answers.

and Development (OECD) was responsible. These answers suggest that a considerable number of participants were not aware of the accountability of people whose data had been exposed in a breach (see Fig. 3). It seems clear that many users either have distorted perceptions of accountability or are unsure what to factor into their attributions of blame in data breach episodes.

In our previous study [10] where we analyzed the nature of a company's communication with consumers regarding data breaches, we found that the company's approach appeared to deflect responsibility by shifting the blame and using compensation strategies that introduce the company as the users' helper to reduce the reputation and financial damage of their data breaches. This finding may influence user understanding of accountability, which indicates the need for a more systematic approach in communicating information on data security to the public.

TABLE III
SAMPLE PARTICIPANTS COMMENTS FOR CAUSES OF A DATA BREACH

Causes	Total	Sample of Data Extracted
Negligence	7	"Carelessness, on the part of the designer, on the user's part for not behaving safely."—P1
Poor password	5	"When people don't use difficult enough passwords"—P30
Poor security measures	8	"A company may cut corners and just do the bare minimum to get certified."—P18
Vulnerability	6	"Vulnerability, which is having data that is not encrypted. The hacker targets vulnerable users more easily"—P3.
Hackers	9	"I believe that an Internet data breach is caused by a hacker"—P25

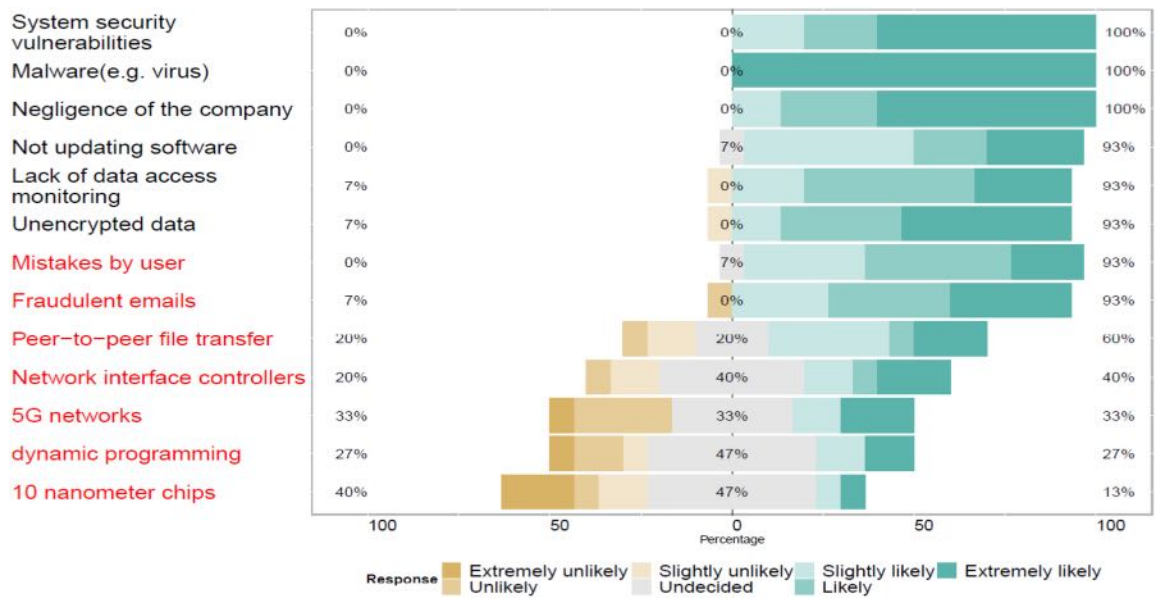


Fig. 4. What causes an internet data breach? Labels in red are inappropriate answers.

2. *Causes of Data Breaches:* We asked questions about causes in both an open-ended and a Likert-scale format. Participants generally mentioned the following items in their open-ended answers: negligence, poor passwords, poor security measures, hackers, and vulnerability. Notably, participants employed the word *vulnerability* in its everyday meaning of "human weakness" rather than its technical use of "system weakness." Other potential causes of data breaches can be seen in Table III. As can be seen in Fig. 4, although the rating seems reasonable for the appropriate answers, more than 40% of our participants were uncertain or rated the inappropriate answers high, indicating that they did not know the causes of data breaches.

3. *Consequences of a Data Breach:* When asked what can happen as a consequence of an internet data breach, participants mentioned, in order of importance, financial loss as the main consequence, identity theft, invasion of privacy, fraud, loss of trust, national threat, political consequences, and cyberwar.

Fig. 5 shows the results from Likert-type questions. Participants rated highly or were uncertain about some inappropriate responses such as data-harvesting programs, browser hijacking, pop-up advertisements, and personal safety. This finding suggests that many nonspecialist users are unfamiliar with the actual consequences of data breaches. Although 47% correctly identified the inappropriate consequences, about 50% still showed no knowledge. This result suggests that if users do not know how stolen information is used, they may not consider the consequences of data breaches as serious.

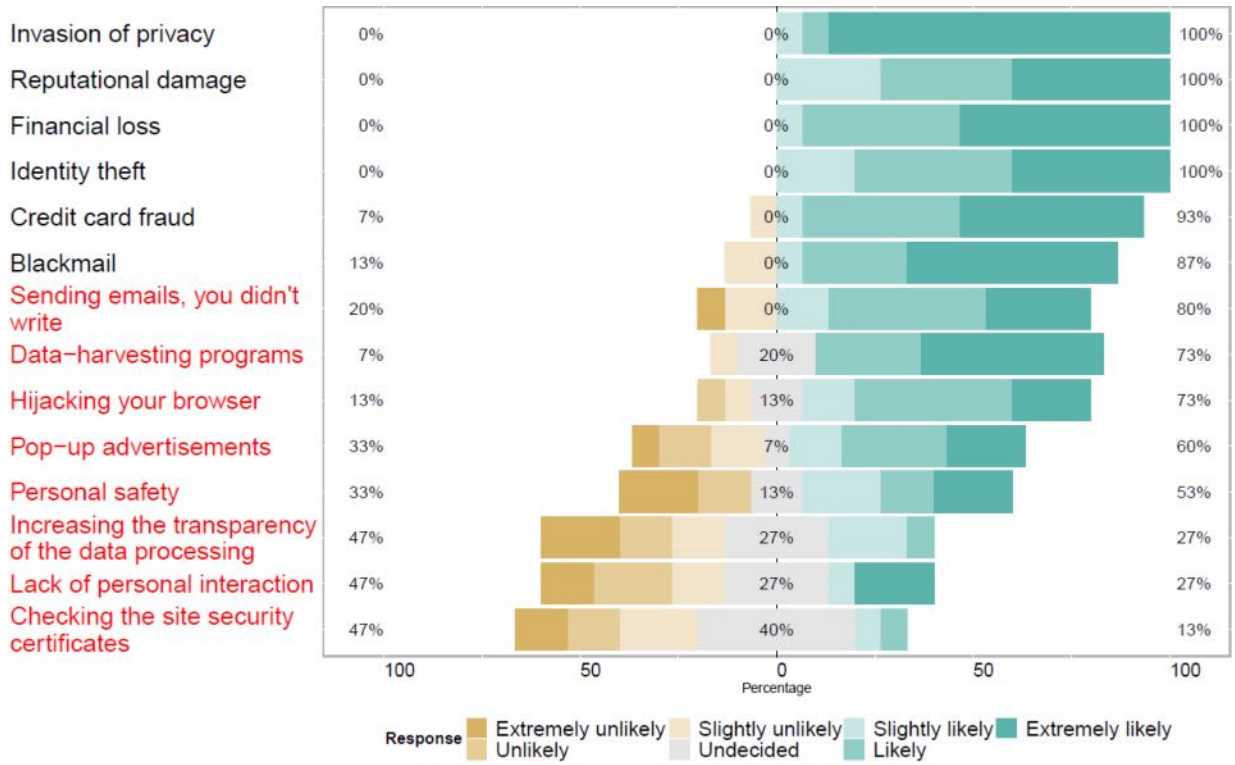


Fig. 5. What causes an internet data breach? Labels in red are inappropriate answers.

4. *Motivation for Data Breaches:* In response to being asked the motivation for an internet data breach, participants mentioned common reasons. Only a few participants had a broader perspective of the factors that can motivate an attacker and mentioned having fun, damaging corporation reputations, spying, proving a weakness in a system, and displaying technical skills.

Participant responses to the Likert-scale questions (see Fig. 6), also suggest that they are familiar with common motivations for data breaches. Surprisingly, however, more than 50% agreed or were uncertain that a data breach can happen to secure electronic data, to stay on top of advancement in security technologies, or to check a company's security measures. These responses suggest an unclear understanding of reasons for data breaches.

5. *Prevention Methods:* When asked what companies should do to protect their data from breaches, participants suggested investing in security measures, educating users and employees, restricting data access and regularly monitoring networks, and encrypting data. Hiring reliable employees and not collecting all data were also considered by one participant as suitable prevention methods. For Likert-scale questions (see Fig. 7), responses were mostly reasonable; however, some responses suggested that participants were simply guessing. For instance, these responses mostly agreed that standardized protocols, high-performance networking, and genetic algorithms can prevent a data breach (all of which show mistaken perceptions). Again, these findings indicate the need for clearer communication in educating users about security.

6. *Actions After a Breach:* In response to being asked what they can do if their personal data have been breached, a few participants talked about freezing their accounts, and only one was aware of an "identity theft protection service." Our results for Likert-scale questions (see Fig. 8) show that many users are only partially aware of appropriate after-breach procedures.

Drawing Task Results Our goal in this part of the study was to explore the users' perceptions of how a data breach occurs. Morgan et al. [12] discuss the importance of eliciting how elements of mental models *influence* other elements—for example, where some events are causally related to other events or outcomes. We asked them to sketch their understanding of the process, and to get them

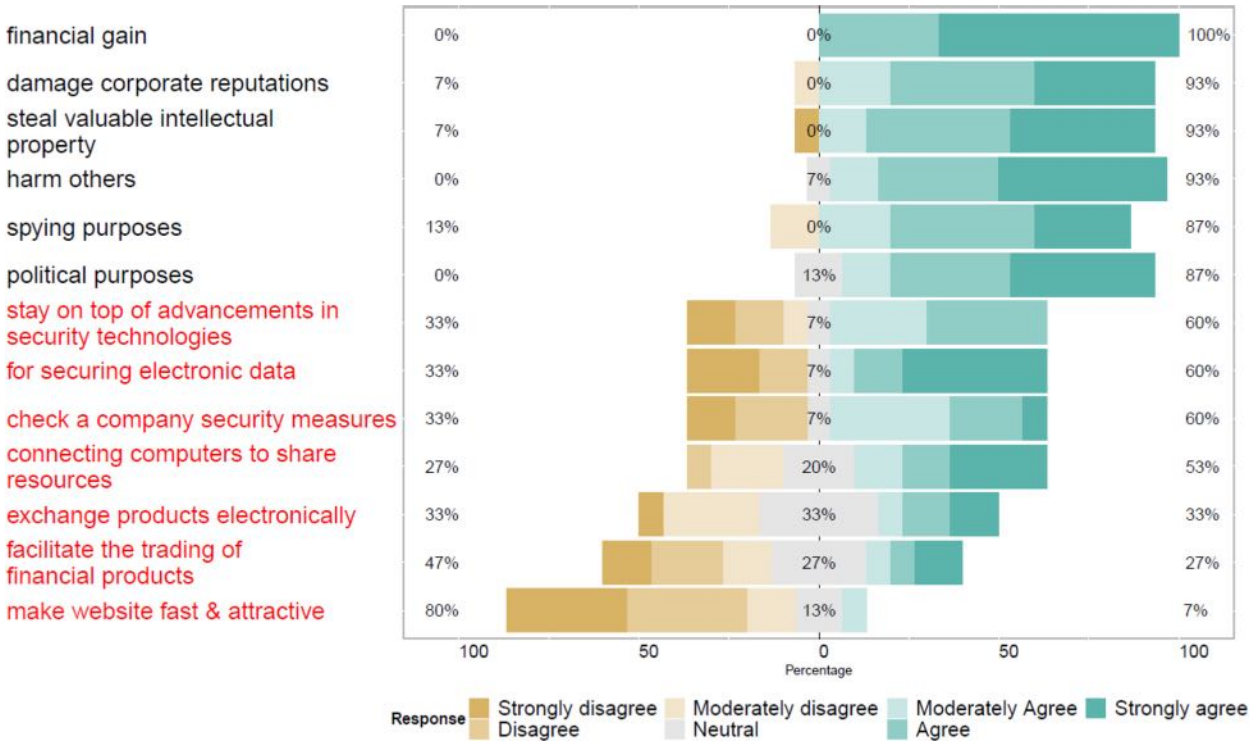


Fig. 6. Why do data breaches happen? Labels in red show misleading perceptions.

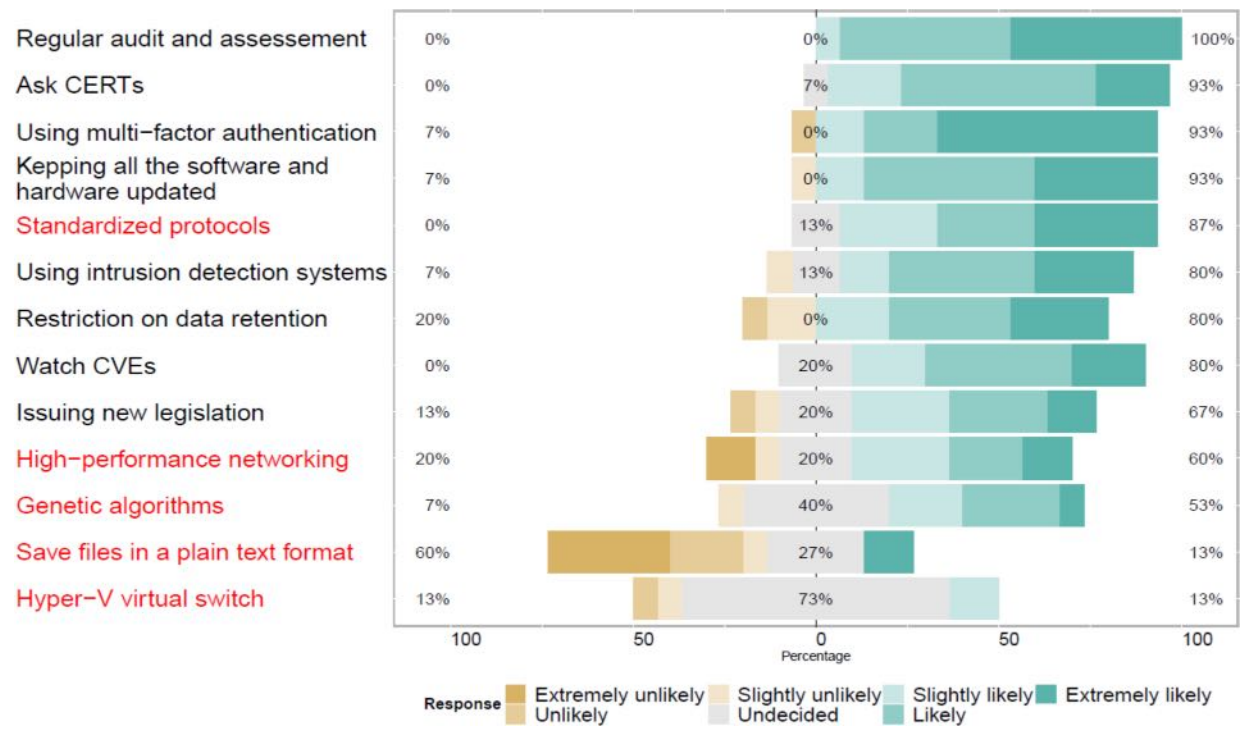


Fig. 7. How likely is it to prevent a data breach via the following practices? Labels in red show misleading perceptions.

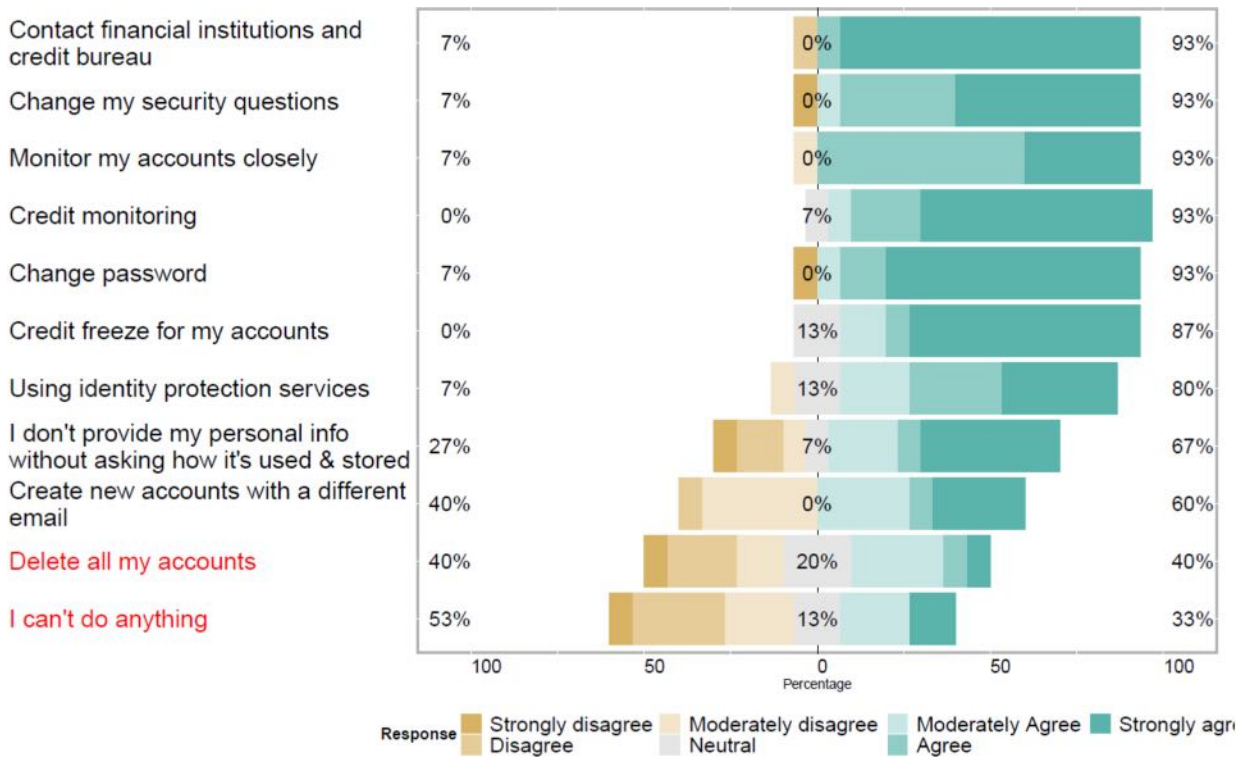


Fig. 8. What can you do if your data has been breached? Labels in red show misleading perceptions.

started, we suggested the following potential elements: Admin, Data, Customer, Hacker, Access to data, and Vulnerability, and asked them to describe what these elements' roles are in a data breach.

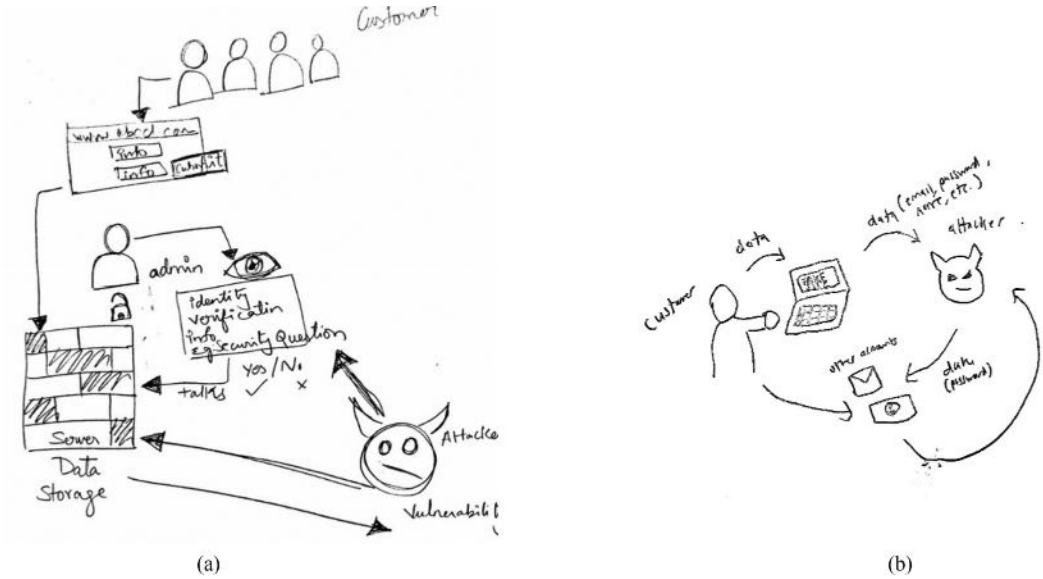
We then conducted a thematic analysis, reviewing the results for each participant and identifying the mechanisms shown that suggested their mental model of the data breach process. Two researchers separately reviewed the diagrams and then discussed their findings to reach agreement on the main themes. The codes identified related the roles involved (e.g., user, administrator, and attacker), the information, location of the attack (e.g., user computer, server, database), and the means by which access was gained (e.g., passwords, special knowledge, vulnerabilities). In some cases, the causal sequence of events depicted needed discussion (e.g., how surveillance found passwords or vulnerabilities). Overall, we identified three major mechanisms in the participants' drawings:

1. That the attacker gains admin access
2. That the attacker gains user access
3. That the attacker gains backdoor access

We also found one more mechanism (database server vulnerability) illustrated by only two participants. The following sections describe these patterns in more detail.

Admin Access: Participants depicted their understanding of how an attacker can gain access to the data as attacking the admin system directly and gaining admin level access to the data. In this mechanism, user data are stored on a server that can be unlocked only by an administrator. These participants also believed that accessing the admin system would happen either by a capture attack on the admin password or by guessing the admin's password. A capture attack is the action of obtaining information such as passwords and other confidential data. It can be accomplished by keyloggers, malicious software or hardware that records keystrokes at a low level and, therefore, records any usernames, passwords, or other credentials entered by typing. Participants were not sure about the level of admin access and one mentioned that the admin may have access only to users' identity verification data, so an attacker who targets the admin account will access those data and will then be able to access the server as a normal user [see Fig. 9(a)].

User Access: This group of participants was focused on *when* attackers try to hack an individual's data. They also believed that it could



(a) (b)

Fig. 9. Participants' drawing 1. (a) P19. Attacker hacks the admin system to steal. (b) P14. Attacker hacks the user system to steal personal data.identity verification data.

happen by password guessing or password capturing attack. Attackers may attempt to take screen captures of a user's personal computer to gather information over the course of an operation or input capture to steal a user's credit-card information while they are shopping online on an apparently legitimate but really fraudulent website [see Fig. 9(b)].

Backdoor Access: In a few cases, participants' drawings of a data breach illustrated an existing backdoor that could be used by attackers. In this case, an attacker sees a protected sign in the frontend that gives access to secured data with the appropriate "key" (customer and admin). Then, the attacker looks for the backdoor and "tries the knob." The backdoor can appear as a result of malware or by an intentional manufacturing decision. However, one participant tried to show that attackers can gain access to the backdoor by brute force attack, checking all possible passwords until the correct one is found. The brute force attack is indeed one actual kind of attack, though other actual kinds of attack include exploiting backdoors that are put in place on purpose by manufacturers or cybercriminals to allow access into a system
[see Fig. 10(a)].

Database Server Vulnerability: In the second round of the study, after changing the drawing prompt, two participants drew an attacker gaining access to the system through database server vulnerabilities. However, they did not show any illustration of the vulnerabilities in their drawings. P27 said that attackers can use different methods to breach customers' data by accessing data through getting administrator credentials, compromising individuals' data by deceiving users to disclose their information (e.g., using phishing attacks), and using database or server vulnerabilities [see Fig. 10(b)].

Based on the participants' drawings, it is not clear whether users understand what a vulnerability in a system is. Three participants (P2, P10, and P28) drew the vulnerability as an open door, and three as a broken wall (P15, P30, and P34), and the others did not include it in their drawings. Only one participant (P13) added more details about the steps of the attack where the attacker scans the system for confidential information. The drawings of two participants (P8 and P17) had labels about the consequences of a data breach, like loss of resources and invasion of privacy.

The main finding of our picture analysis (in the early round) was that most participants believe that attackers are interested in personal data that they have stored on their devices or shared unintentionally with an unauthorized party. For example, P8 said,

> A data breach happens when a stranger gets access to my private pages such as Instagram and … finds my username and password then uploads pictures and information instead of me.

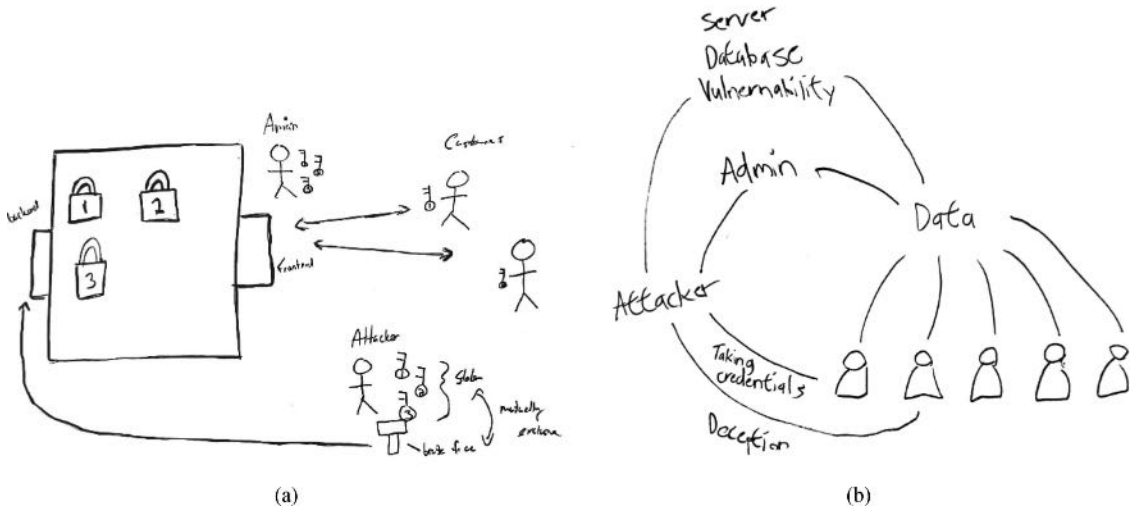


Fig. 10. Participants' drawing 2. (a) P5. Backdoor access. (b) P27. Different types of attack.

P20 said that "A data breach happens when a third party is monitoring your conversations with a server and recording your data without your authorization.

Using elements like customer and data admin did not change their model of how their personal data are disclosed. Even in the second round of study, after we changed the wording of our drawing task prompt to explicitly state the mass consequences of data breaches, some participants still believed that attackers are interested in accessing individual users' credentials.

Discussion

The extracted themes from participants' drawings and other remarks they made suggest how a data breach is represented in their mental models. We now summarize our key insights and discuss their implications for communicating security procedures to users.

Drawing and Open Questions Analysis

Target of Attack: We found that most participants believed that attackers attempt to access their personal information through their social media accounts, emails, or data stored on their personal devices like mobile phones or tablets. During the drawing task, some people illustrated data on their devices and ignored data stored on organizational databases.

Due to this perception of data storage, they believed that because users are entering their data, *they* are the principal people responsible, when it is not true. For example, when asked about data protection responsibility, P8 said, "The most responsible person is us; then it goes to companies we share our information with."

After we changed the wording of our drawing task, they were still uncertain about the attack's target. Most participants still believed that attackers are interested in users (vulnerable users), while fewer included the system admin, and even fewer identified the vulnerability of the system itself as a potential target. This may be the reason why many people continue using breached services.

Uncertainty: Participants expressed much uncertainty about how an attacker gains access to data, and their answers were mostly focused on either guessing passwords or capturing the users' credentials while they are using an illegitimate website, especially for online shopping. Another area of uncertainty in people's knowledge was whether protective actions, like encryption, are strong enough to prevent data breaches. We did not ask directly about the strength of encryption algorithms; however, a few participants talked about encryption as a preventive method. Moreover, although they agreed in Likert-scale responses that encryption is a preventive method, they noted that some of the breached companies had used data encryption that had proven to be ineffective. Our findings are consistent with what Wu et al. [23] found regarding the participants' confusion about encryption strength. This result shows the need to communicate to users the significance of encryption's protective capabilities

so that they may demand that companies encrypt their data using strong encryption algorithms.

Participants know the term *vulnerability*, but they typically think that it relates to themselves. They used *vulnerability* in the sense that users are more vulnerable to attack because of a lack of security knowledge or lack of malware protection. Participants said, "The hacker targets vulnerable users more easily," and "when there is no protection against malware and/or hackers, then the data would be left vulnerable and could potentially be accessed by the attacker." Moreover, only six participants illustrated *vulnerability* as a broken wall or open door in their drawings. This finding highlights clearly the need to define terms to users rather than assuming understanding, especially with terms that have multiple meanings.

Some participants mentioned that users need to know more about the type of data stored on companies' databases and the ways that these data are protected. This finding underscores the need for information flows to be more transparent, for example, by making GDPR compliance clear [29] by stating specifically the purpose of data collection, time period of data storage, permission to access data, and automated decision-making like profiling. This goal can be achieved by improving software design to support users' ability to track access to their data and perhaps giving a user warning when someone accesses their personal data.

Our findings also demonstrate participants' inadequate knowledge of what to do when their data are breached. They did not mention any protection mechanism against the negative impact of data breaches, and it is not clear whether they know that these mechanisms exist. They mentioned changing passwords, deleting accounts, or reporting the incident after-breach.

> Call the police or email/report this incident to the company/Admin or any regulatory body such as investigators.
>
> Talk to police, bank, ISP, depending on the extent.
>
> Change my passwords, delete accounts, and create new ones with a different email.

This finding underscores the need for clear communication on what actions are required after a breach and what can go wrong if these actions are not taken. The study identified a need for information on how to reduce risk of a data breach, for example, by placing a credit freeze, and how to respond to a breach depending on the type of information that was compromised (e.g., type of identity protection services that they need to use).

Questionnaire Analysis The results of the questionnaire show that users were exhibiting safe security behavior in securing their devices, but the scores are lower in updating, proactive awareness, and password choice. Participants correctly identified users' and employees' negligence, weak passwords, and inadequate security measures as causes of a data breach. They understood that financial loss, identity theft, and invasion of privacy are negative impacts of a data breach. A variety of motivations for the attack was mentioned in open-ended questions—for example, attackers displaying their technical skills and political motivations.

Participants tended to rate legislation as less likely to prevent a data breach, and they incorrectly identified answers that we offered such as high-performance networking, and genetic algorithms as prevention methods. This response might be a result of their unfamiliarity with new laws that can force companies to invest more in their security measures but could also underscore the common belief that attackers are competent enough to find a way to penetrate a system. This finding shows an opportunity to communicate to users that many massive data breaches could have been prevented by simple security practices like updating certificates and patching software promptly [4].

Limitations Our work has certain limitations. First, our sample cannot support conclusions about the general population because it is skewed to younger people with a higher level of education, and it is possible that this fact influenced our findings.

Second, although we explored participants' mental models about data breaches, we did not explore the reasons for these mental models. Our earlier study [10] did establish a link between messaging and understanding of issues of responsibility for data breaches, and this topic deserves further exploration.

Conclusion

Users conceptualize security threats, and they create mental models based on their individual experiences. These models shape their behavior

and guide their security-related decisions. In this article, we have presented our findings drawn from 35 sessions with participants about their perceptions of data breaches.

Results from the questionnaire and the drawing tasks suggested that users have a reasonable basic model of how data breaches happen. Participants showed limited detail in their drawings, and they also illustrated that they are not aware of potential causes of a breach like vulnerabilities in a system that can be prevented remedied.

An important finding in our early round was user perception of the responsibility for data protection. Under the European Union's General Data Protection Regulation regulations, Canada's Protection of Personal Information and Electronic Documents Act, and US Federal Trade Commission restrictions, it is businesses' responsibility to keep personal data safe and protected against unauthorized access. However, many participants attributed blame to their own actions and continued to utilize the services of breached companies.

We also found that users' mental models of data access coincided with the category of network attacks against confidentiality [30]. Users understood that an attacker can capture their credentials using screen capturing or input capturing. Participants had a flawed model of what a vulnerability is, and they referred to the user who can be vulnerable due to lack of awareness and knowledge; only a few of them had an understanding of vulnerability in a system and only two participants illustrated database servers' vulnerability as a target of attack. Participant responses for necessary after-breach actions made it evident that they were concerned about checking their bank transactions and changing their passwords. Although identity crime (for example, identity theft, scam, fraud) is one of the most important potential consequences of a data breach, our participants did not seem to know any way to monitor or stop it [2].

In all, the results of this study indicate that many users are not aware of effective actions to prevent data breaches or of actions to manage them when they do occur. As our discussion showed, this is a communication problem, as no systematic or organized effort seems to exist for instructing users on the pertinent factors in data breach crises based on their existing knowledge and mental models. This gap in user understanding is not only costly to the public but also enables organizations to deflect responsibility from their actions and evade public scrutiny through strategies that have been developed to help them to do so. This study proposes and illustrates a mental models approach for obtaining data from users to design targeted communication protocols that will increase awareness of cybersecurity issues. It is intended as a foundation for further research into user perceptions of data breaches, employing different samples and scenarios, and to better identify areas of confusion, which would confirm where focused communication design is required.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2016-06149.

References

[1] I. Goddijn and J. Kouns. (2020 Jul.). Risk based security. 2020 q1 report data breach quick view. [Online]. Available: https://pages.riskbasedsecurity.com/en/2020-q1-data-breach-quickview-report
[2] Identity Theft Resource Center. (2019). [Online]. Available: https://www.idtheftcenter.org/data-breaches/
[3] A. Lazarevic, J. Kumar, and V. Srivatava, "Intrusion detection: A survey," in *Managing Cyber Threats: Issues, Approaches, and Challenges*. Boston, MA, USA, Springer, 2005, pp. 19–78.
[4] GAO-18-559. (2018). Data protection: Actions taken by equifax and federal agencies in response to the 2017 breach GAO-18-559. [Online]. Available: https://www.gao.gov/assets/gao-18-559.pdf
[5] D. Swinhoe. (2021 Jan./Aug.). The 15 biggest data breaches of the 21st century. [Online]. Available: https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html
[6] C. Mele, "Data breaches keep happening. So why don't you do something?," *The New York Times*, Jan./Aug. 2018. [Online]. Available: https://www.nytimes.com/2018/08/01/technology/data-breaches.html
[7] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the gap in computer security warnings: A mental model approach," *IEEE Secur. Privacy,* vol. 9, no. 2, pp. 18–26, Mar./Apr. 2011.
[8] R. Wash, "Folk models of home computer security," in *Proc. 6th Symp. Usable Privacy Secur.*, 2010, pp. 1–16.
[9] R. Wash and E. Rader, "Influencing mental models of security: A research agenda," in *Proc. New Secur. Paradigms Workshop*, 2011, pp. 57–66.

[10] Z. Hassanzadeh, S. Marsen, and R. Biddle, "We're here to help: Crisis communication and user perception of data breaches," in *Proc. Graphic Interface*, 2020, pp. 236–245.
[11] T. Van Djik, *Cognitive Context Models and Discourse*. M. I. Stamenov Ed. Amsterdam, The Netherlands: John Benjamins Publishing, 1997.
[12] M. Granger Morgan, B. Fischhoff, A. Bostrom, and C. J. Atman, *Risk Communication: A Mental Models Approach*. Cambridge, UK: Cambridge Univ. Press, 2001.
[13] P. N. Johnson-Laird, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, UK: Cambridge Univ. Press, 1990.
[14] L. J. Camp, "Mental models of privacy and security," *IEEE Technol. Soc. Mag.*, vol. 28, no. 3, pp. 37–46, Fall 2009.
[15] P. N. Johnson-Laird, "Mental models and human reasoning," *Proc. Nat. Acad. Sci.*, vol. 107, no. 43, pp. 18243–18250, 2010.
[16] M. T. H. Chi, "Laboratory methods for assessing experts' and novices' knowledge," in *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK, Cambridge Univ. Press, 2006, pp. 167–184.
[17] D. Jonassen and Y. H. Cho, "Externalizing mental models with mindtools," in *Understanding Models for Learning and Instruction*. Boston, MA, USA: Springer, 2008, pp. 145–159.
[18] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental models: An interdisciplinary synthesis of theory and methods," *Ecol. Soc.,* vol. 16, no. 1, pp. 46–58, 2011.
[19] M. T. H. Chi, P. J. Feltovich, and R. Glaser, "Categorization and representation of physics problems by experts and novices," *Cognit. Sci.*, vol. 5, no. 2, pp. 121–152, 1981.
[20] D. H. Jonassen, "Computers as mindtools for schools: Engaging critical thinking," *J. Educ. Comput. Res.*, vol. 23, no. 2, pp. 233–235, 2000.
[21] D. H. Jonassen, "On the role of concepts in learning and instructional design," *Educ. Technol. Res. Develop.*, vol. 2, no. 54, 2006, Art. no. 177.
[22] F. Raja, K. Hawkey, and K. Beznosov, "Revealing hidden context: Improving mental models of personal firewall users," in *Proc. 5th Symp. Usable Privacy Secur.*, 2009, pp. 1–12.
[23] J. Wu and D. Zappala, "When is a tree really a truck? Exploring mental models of encryption," in *Proc. 14th Symp. Usable Privacy Secur.*, 2018, pp. 395–409.
[24] F. Asgharpour, D. Liu, and L. J. Camp, "Mental models of security risks," in *Financial Cryptography and Data Security*, vol. 4886. Berlin, Germany: Springer, 2007, pp. 367–377.
[25] S. Egelman and E. Peer, "Scaling the security wall: Developing a security behavior intentions scale (SeBIS)," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 2873–2882.
[26] G. Guest, K. M. MacQueen, and E. E. Namey, "Introduction to applied thematic analysis," in *Applied Thematic Analysis*. Thousand Oaks, CA, USA: Sage, 2012.
[27] J. Check and R. K. Schutt, "Qualitative data analysis," in *Research Methods in Education*. Kent, UK: Croom Helm, 2012.
[28] U. Flick, "The SAGE handbook of qualitative data analysis," in *Research Methods in Education*. Thousand Oaks, CA, USA: Sage, 2014.
[29] R. Koch. (2018). Everything you need to know about GDPR compliance. [Online]. Available: https://gdpr.eu/compliance/
[30] I. Sumra, H. Bin Hasbullah, and J. L. Ab Manan, "Attacks on security goals (confidentiality, integrity, availability) in VANET: A survey," in *Vehicular Ad-hoc Networks For Smart Cities*, vol. 306. A. Laouiti, A. Qayyum, M. S. Mohamad, Eds. Singapore: Springer, 2014, pp. 51–61.

Zahra Hassanzadeh received the master's degree in computer science from Carleton University, Ottawa, ON, Canada, in 2019. She is currently a User Experience Researcher with Public Service and Procurement Canada in Ottawa, Ottawa, ON, Canada. She also has an educational background in translation and applied linguistics. Her research is focused on usable security and user-centered design to create systems and services that fit better into peoples' daily lives.

Robert Biddle is currently a Professor with Carleton University, Ottawa, ON, Canada, appointed both to the School of Computer Science and the Institute of Cognitive Science. His research is primarily in human factors in cybersecurity and software design, especially creating and evaluating innovative designs for computer security software and collaborative software development.

Sky Marsen is currently the Director of the Media and Communication Program with Flinders University, Adelaide, SA, Australia. She has an educational background in applied linguistics, semiotics, and communication studies, and has lectured in these subjects internationally. She has published widely on her research interests, which include crisis communication, organizational communication, narrative theory, and discourse analysis. Dr. Marsen is an Associate Editor for the *International Journal of Business Communication*.

Research Article

Characterizing Disciplinarity and Conventions in Engineering Resume Profiles

—Catherine G. P. Berdanier, Mary McCall, and Gracemarie Mike Fillenwarth

***Abstract*—Background:** *Resume preparation is a common activity within technical writing classes, but the advent and increased use of resume profile and job-hunting sites, such as Indeed.com, require instructors and researchers to re-think common practices in the teaching of resume writing, particularly for writing instructors with limited disciplinary experience. Prior research for conventional resumes has quantified the disciplinarity of resumes as a function of resume quality using metrics of disciplinary discourse density, which may be useful in analyzing online resumes profiles.* **Research questions:** *1. How do online engineering resume profiles demonstrate disciplinarity? 2. What formatting and stylistic conventions are observed within engineering resume profiles? 3. How do rhetorical disciplinarity and conventions vary with resume profile quality?* **Literature review:** *Although past efforts have examined the resume as a critical genre for entering a professional setting, few researchers have sought to interpret the relationships between discursive and stylistic expectations and quality in online resume profiles, while also accounting for aspects of disciplinarity.* **Methodology:** *This study compares engineering (all disciplines) resume profiles from Indeed.com with a corpus of conventional engineering resumes through qualitative genre analysis and quantitative methods for calculating disciplinary discourse density. We also characterize stylistic and rhetorical conventions for resume profiles, and statistically compare these facets as a function of resume quality.* **Results and conclusion:** *Results determined that discursive strategies were significantly different between strong, moderate, and weak engineering resume profiles. Qualitative analysis captured differences in style and form that were also statistically linked with quality. Based on our results, we call for further investigation into resume profiles and reconsideration of current pedagogical approaches.*

Index Terms*—Communication, disciplinary discourse, engineering resumes, professional development, resume profiles.*

There has been a substantial shift in hiring practices over the past decade toward electronic job search, recruitment, and application tools. Adding one's resume to individual employer sites and jobs sites, such as Indeed.com and Monster.com, is becoming an increasingly common, if not standard, practice. While some online networking and job search sites enable candidates to upload a conventional resume in Word or PDF format, other sites, such as Indeed.com and LinkedIn, ask applicants to submit their qualifications through prebuilt forms. Despite this shift, the conventional print resume holds the focus of most technical writing textbooks and web resources aimed at job documents, leaving instructors with few resources on preparing students for the modern job application process.

In addition to the need to consider evolving technologies for the teaching of resumes, instructors who teach resume writing also face the challenge of each discipline having some distinct conventions as well as some overlapping resume-writing conventions. At the same time, students often find resume advice in their textbooks to be outdated, generic, and irrelevant to their field [1], making the teaching of resumes a uniquely challenging, high-stakes task. While navigating these issues, resume instructors and writers must also consider other factors concerning the ways that resumes will be read and used, considering rapidly evolving technologies and disciplinary norms, such as the emergence of resume profiles. Given these shifts, the validity of the "rules" for resume writing as they have been traditionally taught warrants additional investigation.

In this article, we use qualitative and quantitative analyses to assess the quality and disciplinarity of online resume profiles and understand the typical

Manuscript received May 9, 2019; revised May 31, 2021; accepted June 1, 2021. Date of publication October 26, 2021; date of current version November 24, 2021. *(Corresponding author: Catherine G. P. Berdanier.)*
Catherine G. P. Berdanier is with the Department of Mechanical Engineering, Penn State University, University Park, PA 16802 USA (email: cgb9@psu.edu).
Mary McCall is with the Department of English, North Dakota State University, Fargo, ND 58108 USA (email: mary.mccall@ndsu.edu).
Gracemarie Mike Fillenwarth is with the Department of Writing Arts, Rowan University, Glassboro, NJ 08028 USA (email: fillenwarth@rowan.edu).

IEEE 10.1109/TPC.2021.3110397

Practitioner Takeaway

- The disciplinarity of engineering resume profiles can be assessed quantitatively using metrics of disciplinary discourse density, a way of attributing numerical values to disciplinary discourse.
- Strong engineering resume profiles (from Indeed.com) differed statistically significantly from weak resume profiles with respect to disciplinary discourse density and use of conventions such as use of bulleted lists or verb-centric phrasing.
- Rather than being taught resume writing as a monolithic genre, students should be taught to demonstrate disciplinary mastery across conventional and emergent activity systems related to career preparation, such as online resume profiles.

style and format conventions used within engineering resume profiles. A variety of terms has been used in the literature to describe various resume formats hosted or uploaded to the web as digital applications evolve over time (e.g., web resumes [2], scannable resumes [3], online resumes [4]–[6], and electronic resumes [7], [8]). However, for the purposes of this article, we define "conventional" resumes as (normatively) one-page documents intended to be submitted to and reviewed by an employer and able to be easily printed out. In contrast, we define online "resume profiles" as the quasiresumes that are native to an online job search website, such as Indeed.com. Rather than being created in a word processing application, these resumes are constructed through online resume builders that guide users in the creation of a resume through a series of predetermined sections while also allowing ample personalization by including an open-ended description section for each experience. As one example, the current version of LinkedIn includes a similar feature that enables users to build their profiles, which may be used for job-searching purposes as well as for the site's larger professional social networking function.

In response to these modern manifestations of the resume and the challenges that they bring for writing instructors, while also anticipating future evolution of the genre, this article posits that disciplinary discourse is a bridge that can help instructors and engineering students alike consider expectations for resumes and resume-like genres in newer formats, such as resume profiles on job-search websites. The purpose of this study is to discern features of quality in engineering resume profiles, relying heavily on disciplinary discourse and disciplinary discourse density, a method proposed in prior work [9]. This research seeks to answer the following questions.

RQ1. How do online engineering resume profiles demonstrate disciplinarity?
RQ2. What formatting and stylistic conventions are observed within engineering resume profiles?
RQ3. How do rhetorical disciplinarity and conventions vary with resume profile quality?

As a result of this study, we call for additional research on resume profiles and the development of teaching methods that help engineering students and instructors navigate emerging genres of job-search materials.

Literature Review

Given the longevity of genres, such as the resume and cover letter within the job-hunting process, research across the disciplines has been invested in best resume-writing (and pedagogical) practices to prepare students for internship opportunities and postgraduate careers. Between the 1970s and 1990s, studies largely examined topics such as the order of information in resumes and comparisons between students, instructors, business communication textbook advice, and recruiters on resume preparation [10]–[17]. Attention to emerging, persuasive approaches to resume content, style, and delivery methods still remains prevalent [18]–[22], but more recent research has focused on comparisons between print and video resumes [23]; students' learning processes while crafting resumes [24], [25]; and the impact of race and gender on resume interpretation [26]–[28]. Other recent, comprehensive literature reviews of resume scholarship include research from disciplines such as career development and applied

psychology [1] and in technical and professional communication and STEM education journals [9].

Emerging digital technologies have ushered in discussions about new formats like the scannable resume and updated resume-writing practices for the web [29]–[32], compelling classroom assignments that guided students on how to create web-based versions of their conventional resumes. A new direction in resume research comes with the prevalence of online job-hunting websites that employ their own platforms for users to input experience data. The text boxes provided on these websites do not require specific formatting, nor are they necessarily intended to be printed out, although that is an option. Further, the job applicant often does not submit a tailored resume profile to a specific company using LinkedIn or Indeed.com; rather, the onus is on the employer to use such sites to "find talent," sometimes using a digital applicant tracking system (ATS) to digitally filter resumes to deposit a set of relevant and qualified candidates to a hiring manager.

Acknowledging the Role of ATSs While acknowledging the ATS as a valuable tool for employers but a potential obstacle for an applicant, articles that do discuss algorithmic filtering systems in sorting resumes and resume profiles acknowledge the need for applicants to learn to write digital resumes that can be applicable to multiple audiences. As a specific example, Diaz's [33] discussion of "electronic" and "scannable" resumes discusses the use of optical character recognition software, a precursor to modern robust text parsing systems that facilitate job application websites like LinkedIn and Indeed.com. She notes that

> updated best practices need to focus on writing one resume without knowing precisely how the employer, or even a specific resume reviewer, will approach the resume

and that

> once an applicant submits a resume, the applicant loses control over what happens to it: A paper resume may become an electronic document read by a computer search engine, and an electronic resume may become a paper resume read by a ballerina doing temp work between productions. [33, p. 433–434]

Most recently, Randazzo's [34] study of resume decisions from the points of view of applicants and employers notes that the employers interviewed said that they interact with resumes and resume profiles at a variety of points in the application process, such that the use of an ATS is common but not universal. Her study also found that although some applicants more specifically chose to design their online profiles and resumes for ATS optimization with respect to rhetorical choices and deliberate use of keywords, employers still expected pragmatic variation in the content, rhetoric, style, and form for resumes, especially for professionals with more expertise, when the profile was reviewed by human decision-makers.

Current advice from Indeed.com [35] notes that approximately 40% of employers use an ATS to prefilter candidates, using both the included information (particularly the inclusion of keywords and keyword synonyms with respect to a job position) and font styles, noting that digital ATSs have issues reading some bullet characters and font families (e.g., serif fonts and nonstandard bullets). Although the inclusion of keywords is important in getting past the ATS and into the hands of a human hiring agent, advice from Indeed.com emphasizes the importance of remembering that the point of the online profile is to demonstrate capabilities to a human employer.

Disciplinarity in Job Documents and the Purpose of the Paper Literature that studies or promotes a disciplinary approach to the study or teaching of resumes is limited, especially in engineering. Disciplinarity does indeed matter in resume writing and evaluation. Charney et al. [36] documented disciplinary differences between business and engineering recruiters' interpretations of resumes. Engineering recruiters, for example, focused heavily on descriptive titles and project design details and methods, compared with business recruiters who focused on the pitch and rationale for the projects described in a resume. These findings motivate a disciplinary approach to the teaching of engineering resume writing.

More recently, attention has turned toward the development of technologies such as machine learning to distill and filter digital resumes from an employer's point of view. For example, Valdez-Almada et al. [37] used machine learning and text mining methods to generate knowledge profiles for software engineers based on resumes to help employers identify technical knowledge in initial digital screening of job applicant resumes. Elliott's recent dissertation [38] similarly used machine learning to mine engineering resumes to

determine engineering competencies and attributes, again with the primary application of the research being to enable an employer to screen applicants quickly. These studies offer little advice to instructors or engineering students, nor do they discuss aspects of stylistic convention or rhetorical decision-making that can be integrated back into the professional communication classroom.

Motivated by this need, our past work examined the role of disciplinary discourse—the words, phrases, or patterns of language that have embedded meanings or values to certain disciplinary communities—in engineering students' and practitioners' conventional resumes to theorize how students can best tailor those resumes for engineering audiences. Through that research, we discovered a link between the quality of a resume and the way that disciplinary discourse is employed: namely, higher quality resumes were characterized by a statistically significantly higher density of engineering lexical occurrences [3], a parameter we called *disciplinary discourse density*. These findings suggested potential interventions in the teaching of resumes that could be introduced not only to engineering students but also to students across various disciplines.

In this article, we acknowledge the role of ATSs as they apply to the process as a whole, but we are interested in characterizing the conventions and disciplinarity present within engineering resume profiles broadly, taking the point of view that our study of disciplinarity and conventions in online engineering resume profiles starts after the resume profile has landed in the hands of a human (e.g., the goal of the study is not to analyze which conventions, style, or disciplinarity get past an ATS). Pulling from Diaz's and Randazzo's recommendations, as well as current advice from Indeed.com on not ignoring the human element while embedding thoughtful and relevant keywords "for" an ATS, we make the assumption that the applicants responsible for the resume profiles in our study understood that their resumes might be filtered digitally before being examined by a human hiring manager or employer, either on a computer screen or printed out for closer review. However, we also assume that the applicants hoped and planned for a human employer to evaluate their resume profile at some stage of the hiring process. From there, our goals in this work are to use quantitative measures of disciplinary discourse density and qualitative comparisons of conventions to propose standards of quality for engineering resume profiles that can be more agile in web contexts.

Theoretical Framework: Activity Theory This study is framed from the point of view of Activity Theory. Derived from Vygotsky [39], [40]; Leont'ev [41], [42]; and later developed by Engeström [43] and Cole and Engeström [44], Activity Theory [also known as cultural-historical activity theory (CHAT), which was introduced by Cole [45] to help unify the developments of activity theory stemming from Vygotsky's work] is concerned with the analysis of the system of "ongoing, object-directed, historically conditioned, dialectically structured, tool-mediated human interaction[s]" [46, p. 510] in which every human behavior is grounded.

When performing activity system analysis, rhetorical genre studies scholars often reference the model depicted by Engeström [43] to identify the various nodes that comprise an activity system [47]. These nodes, which include the "subjects, "mediational means," and "objects/motives," interact inseparably with one another to produce outcomes; these interactions, in turn, are supported by "rules/norms," "community," and "division of labor" [48]. Paraphrasing Kaptelinin [49], Spinuzzi [50, p. 450] identifies the object node at the center of the activity system, describing it as "the 'sense-maker' around which the rest of the unit of analysis, the activity system, forms." This object guides the action(s) of the subjects using mediational means such as physical tools (e.g., computer) or discursive tools (e.g., genre) to achieve the outcome(s). In the creation of the object, the subject may draw from "rules" established by the community concerned with the transformation of the object, and other agents involved in the activity—i.e., the division of labor—may influence the development of the object. More of "a powerful and clarifying descriptive tool" rather than "a strongly predictive theory" [51, p. 7], activity theory is a lens for understanding how such interactions within a system are mediated by the use of tools and symbols (e.g., language).

Activity theory has two main benefits for the study of resumes. First, it invites a deeper analysis of the development and impact of this genre in relation to the activity of resume writing for job search purposes in a field. Second, it affords the opportunity to consider not only the individual subjects and agents involved in the production of a resume, but also the specific disciplinary community and its coinciding rules, tools, and division of labor that influence the creation of the

TABLE I
ACTIVITY THEORY NODES FOR CONVENTIONAL RESUMES AND ONLINE RESUME PROFILES

Activity Theory Nodes	Conventional Resumes	Online Resume Profiles
Artifacts/Tools	Computer, Word Processing Software	Computer, Internet, Platform-specific Profile Builder
Rules	Typical "rules" for formatting (e.g., bullets, verb-focused statements, 1 page)	Few defined guidelines: some platforms may have character/word limits, or suggest guidelines for readability by digital systems if employers choose to abide by them
Object	Conventional 1-page resume, intended to be viewed on a computer or printed as a handout	Resume profile is housed online. Human recruiters will download resume profiles and/or view them online, and print them if desired.
Division of Labor	Applicant expected to tailor resume to each specific company and position	Applicant creates one resume profile hosted online to meet multiple potential jobs; employer does resume filtering work, potentially with the help of an ATS
Subject	Engineering Job Applicant	
Community	Engineering Disciplinary Community	

resume and its effectiveness. The ability of activity theory to "illustrate the dialectical relationship between genres, individuals, activities, and contexts" [52, pp. 98, 102] makes it particularly suited to contextually tracing the development of the artifact of resumes as subjects (resume writers, employers), media (print, digital, online, web), rules (genre conventions), communities (disciplines, employment websites), and divisions of labor (employers, applicants) continually interact.

We draw on Spinuzzi's [50] approach to methodologically and theoretically contracting an object that emerges from his analysis of how the object has expanded over time. Spinuzzi categorizes this historical development as four method-movements: the triangular representation of an activity system as developed by Engeström [43], the concepts that two activity systems can share an object and that networks of activity systems may share components, and the transformation of Engeström's [53] conception of multiperspectival "runaway objects" via "knotworking" or irregular, distributed collaborations [54] within *mycorrhizae* or horizontal connections across activity systems. Applying this framework to our research, we analyze the production of an engineering resume and resume profile as the object(ive) that defines the activity of seeking employment opportunities.

At first glance, it may seem that conventional resumes and resume profiles operate within the same activity system, given their shared purpose. Both documents are used to apply for jobs in an industry or disciplinary context (in our case, engineering). When printed, the online resume profiles look much like a conventional resume. However, deeper analysis shows that the resume profiles operate in a parallel but nonequivalent activity system. Table I shows the differences. Although the subject (the engineering job applicant) and the community (the engineering disciplinary community) are still the same, the other nodes of the activity system are quite different.

The only overlapping connection between nodes is the subject–community junction. We posit that this junction is mediated by the ways that the subject communicates value, expertise, and experience with the wider community using discipline-specific language. Because of this link, we aim to characterize the disciplinarity of a corpus of resume profiles using our previous method for quantifying disciplinary discourse density to begin to discern patterns of disciplinarity. Further, because the "rules" for online resumes are ill-defined, this study begins to benchmark current conventions present for online engineering resume profiles.

Understanding that each activity system is a situated system within a larger network of related subjects, objects, artifacts, and rules dependent on the culture of the discipline or even a particular company (such as whether a company uses an ATS to pre-filter resumes), we must bound our activity

system. In this study, we define our activity system to *start after the potential use of an ATS*, when the resume is back in human hands, such that the ATS is not a relevant node in our interpretation of the activity. The ATS is instead embedded into the "division of labor" node to reflect that a human actor chose to use an ATS and defines keywords. After the ATS prescreens applicants, the engineering resume profiles are evaluated by the human employer.

We posit that this decision is appropriate for several reasons. First, although the authors of the resume profiles may or may not have considered the role of the ATS in their resume design, they certainly anticipated that their resume profile would be evaluated at some point by a human who would expect engineering disciplinary expertise. Further, deep disciplinary expertise will extend far beyond the keywords defined by an employer. Indeed, Randazzo [34] found that the depth of an applicant's experience—in addition to those experiences and competencies directly synonymous with keywords—is important to employers, especially for well-established candidates, regardless of whether the employers have chosen to use an ATS for preliminary screening. We follow the approach of Randazzo [34] to acknowledge the possible use of an ATS but to focus instead on the characterization of the genre of resumes and resume profiles from the point of view of the human hiring manager or employer. For these reasons, examining quality, disciplinarity, and conventions in engineering resume profiles from this defined activity system is valuable.

Research Methodology

This research employs both qualitative and quantitative analysis techniques to quantify rhetorical patterns in resume profiles and to qualitatively assess common conventions in engineering resume profiles across disciplines. The disciplinary discourse density scores and the conventions are then analyzed to understand how these parameters correspond with the quality of engineering resume profiles. In this section, we discuss methodological details on sampling, cleaning of the data, and data analysis. The research team includes engineering and communication researchers who together hold advanced degrees in engineering, English, and engineering education. Our methodological development is consistent with value to the technical communication community, aligned with engineering disciplinary expectations and values, and respectful of rigorous qualitative and quantitative research traditions.

Sampling and Description of the Corpus We compiled a corpus of $N = 200$ engineering resumes available publicly on the job-hunting and recruiting website Indeed.com from a variety of experience levels for all disciplines of engineering. Because of the public availability of the resumes, this research was classified as exempt and did not require Institutional Review Board approval. No efforts were made to quota sample based on engineering field, gender, or final academic degree, though we sought individuals who earned at least a bachelor's degree in an engineering discipline, engineering technology, computer science, or very closely related science discipline.

Resumes of individuals who held engineering degrees but were seeking jobs in other areas, such as business or sales, were removed from the corpus to maintain focus on the ways engineers communicate their expertise to other engineers. Several of the removed profiles applied "engineer" to other professions, such as audio engineers who work in recording studios, and Information Technology (IT) and network engineers, whose degrees in IT are applied to the administration of computers, printers, and company networks. A log was kept for consistency and justification for the eliminations from the data set.

After cleaning the corpus in these ways, a total of $N = 89$ engineering online resume profiles remained for analysis. All participants represented by the publicly available corpus were applying to work in the US, and all resumes were in English. No effort was made to assume gender or racial ethic background because we mined the resumes from a publicly available site and could not ask participants to self-identify these personal data. Resume profiles were downloaded in PDF format using Indeed.com's digital tools, just as a potential employer might download a resume to save it for reference.

Quantitative Resume Analysis Methods employed to analyze the resume profile follow those that we developed for earlier studies of conventional resumes [9], [55], [56], but are described again briefly here. The method is intended to quantify the disciplinary discourse density within engineering resumes by ranking disciplinary words and phrases (lexico-grammatical features) at higher or lower levels of disciplinarity. Coding is based on the engineering competency model first proposed by

TABLE II
ENGINEERING COMPETENCY TIERS BASED ON THE FRAMEWORK FROM THE AMERICAN ASSOCIATION OF ENGINEERING SOCIETIES [57] WITH CORRESPONDING NUMERICAL SCORE

	Example Competencies (Non-Exhaustive)	Numerical Score
Tier 1: Personal Competencies	Interpersonal skills; integrity; professionalism; initiative; dependability; reliability; adaptability; flexibility; lifelong learning	1
Tier 2: Academic Competencies	Mathematics; science and technology; communication (verbal, written, visual); critical and analytical thinking; basic computer skills; school-related research skills	2
Tier 3: Workplace Competencies	Teamwork; client/stakeholder focus; planning; organizing; creative thinking; problem-solving; decision-making; seeking and developing opportunities and solutions; working with tools and technology (e.g., engineering software packages); business fundamentals; teaching (not as professor)	3
Tier 4: Industry-Wide Technical Competencies	Foundations of engineering; design; manufacturing and construction; operations and maintenance; ethics; business, legal and public policy; sustainability and societal/environmental impact; engineering economics; quality control and quality assurance; safety; health; security and environment; general research competence; ability to write grants; publish internal reports; global competency	4
Tier 5: Industry/Sector Functional Areas	Demonstration of specialized expertise; industry-specific research; teaching at university level as an expert; earned advanced degrees; obtain industry-specific funding; membership in professional societies; connections to research advisor	5
Tier 6: Management, Technical Leadership and Occupation-Specific Competencies	Occupation specific requirements; management competencies; staffing; informing; delegating; networking; monitoring work; entrepreneurship; supporting others; mentoring; strategic planning and action; preparing/evaluating budgets; developing an organizational vision; monitoring and controlling resources	6

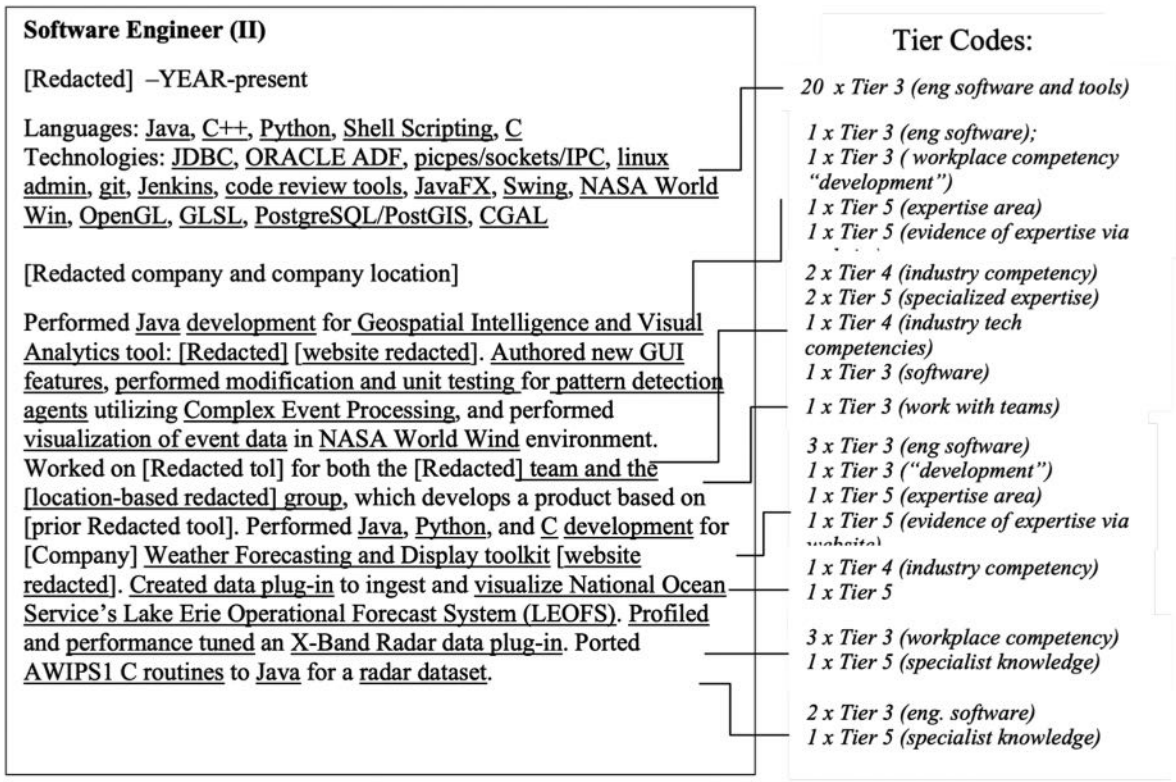
the American Association of Engineering Societies (AAES) [57] and documented in our prior work [55], [56] (note that the AAES has since dissolved). This six-tier model shows various "levels" of competency, with level one skills being professional skills like professionalism and integrity, and levels 5 and 6 skills being demonstrations of expertise and technical leadership within engineering. The framework is shown in Table II.

Words and phrases served as the unit of analysis such that the AAES model acted as an *a priori* framework by which to qualitatively code the resumes. Each coded phrase was assigned the number of the representative tier. Since more sophisticated and advanced engineering-specific competencies are on higher tiers, the coding schema serves as a quantitative measure of the level of disciplinarity within a resume. In coding, attention was paid to code only the rhetorical phrases given, avoiding inferences about roles, responsibilities, and scope unless provided within the coded phrase. Within a phrase, careful attention was given to the assigned value to the tier score of both the action and the demonstration of specialty areas because the resume profiles described roles, competencies, subject matter expertise, and experience with specialized software or applications to specific domains.

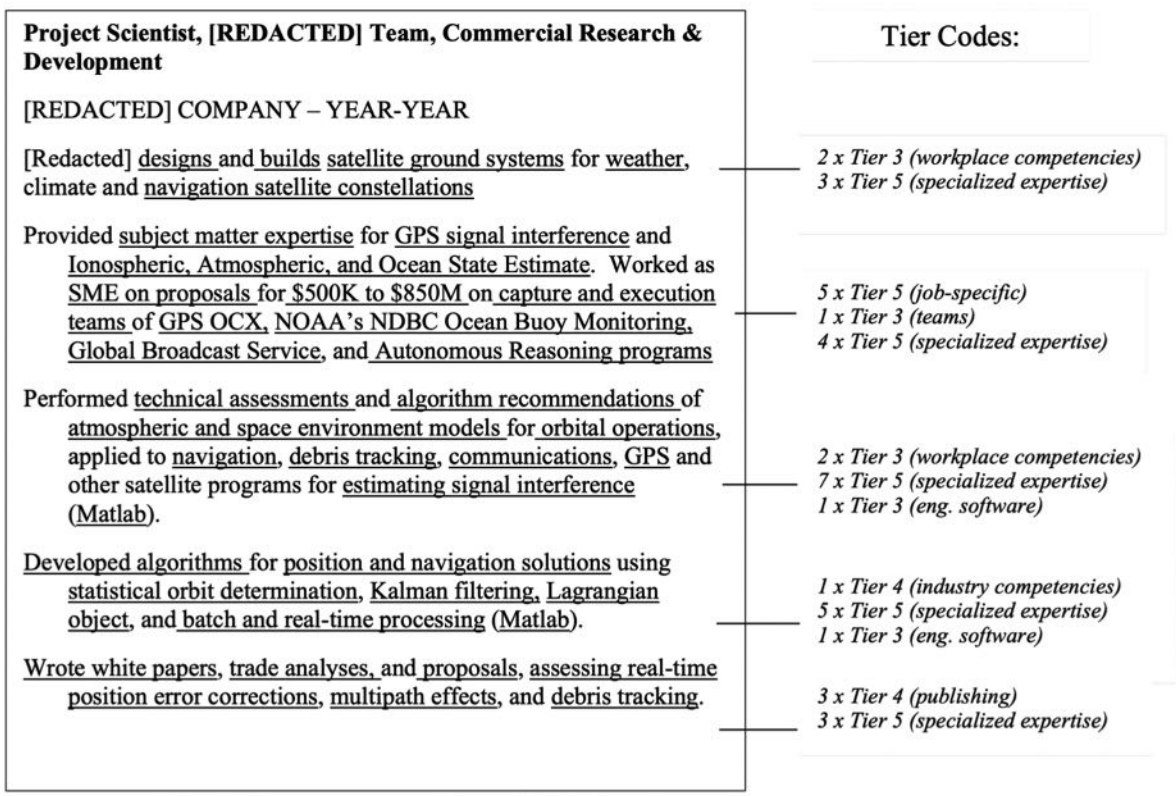
Examples of the coding schema as applied to resume excerpts are shown in Fig. 1. We calculated the disciplinary discourse scores (equation 2) for each document divided by the number of coded phrases to give the "overall disciplinary discourse density" for each resume

$$\text{Sum of Tier Scores} = \sum \text{Code Tier Score} * \#\text{ of Code Occurences} \quad (1)$$

$$\text{Disciplinary Discourse Density} = \frac{\text{Sum of Tier Scores}}{\text{Total Number of Codes in the Resume}}. \quad (2)$$



(a)



(b)

Fig. 1. Demonstration of coding schema for disciplinary discourse density. (a) Excerpt from a "strong" software engineering resume profile. (b) Excerpt from a "strong" aerospace engineering resume profile.

Evaluating Quality of Resume Profiles There are no existing rubrics available to assess the quality of resume profiles that result from online platforms such as Indeed.com. Therefore, we adapted an existing rubric, originally developed by the University of Iowa [58], that we employed in our previous work [9], [55], [56] to sort resume profiles into strong, moderate, and weak categories. This rubric was originally selected because it was one of the few rubrics available that explicitly includes the use of engineering-specific language in the criteria.

Many stylistic conventions "required" for conventional resumes appeared to be non-normative among the resume profiles. The largest example of this discrepancy is the individuality allowed in design: Conventional resumes are typically personalized to the writer's style and can be modified in terms of the use of columns, different fonts, novel section headers, and the like. Resume profiles, however, did not deviate from the prescribed format imposed by the online interface into which users type or copy and paste their relevant career information. Further, traditional conventions such as the "one-page" limit, consistent use of bullets, and short, verb-centric phrases were not regularly employed in the resume profiles.

To methodically investigate these characteristics and to understand how these characteristics varied across quality groups, we removed these stylistic aspects in our revised rubric for resume profiles to avoid self-fulfilling hypotheses. The revised rubric (see Fig. 2) delineates criteria that represent varying levels of quality (3 = strong, 2 = moderate, 1 = weak) across a variety of different aspects of resume writing to capture quality while removing features unique to conventional resumes. Resumes that scored mostly 3s across the criteria were considered strong, and so on. Two of us conducted the quality characterization to the agreement.

These measures of quality are used to statistically compare the differences in rhetorical and stylistic conventions between the quality categories in resume profiles. First, we sorted the resumes into quality groups (strong, moderate, weak) based on the rubric. Then, we analyzed the resumes qualitatively through a genre analysis to calculate disciplinary discourse density scores for all the resumes. Next, we systematically collected data on page length, the extent to which the resume used bullets (which we will refer to as "format"), and the extent to which the resume employed verb-oriented phrases (referred to as "style"). Finally, we conducted quantitative analyses [Analysis of Variance (ANOVA) and chi-square analyses] of these results to discern whether there are statistically significant differences between quality groups with respect to these characteristics.

Results

The $N = 89$ resume profiles analyzed in this study were distributed among the categories of strong, moderate, and weak, as determined by the rubric analysis portion of the analysis: 22 profiles were characterized as strong, 34 as moderate, and 33 as weak. First, we demonstrated and discussed the disciplinary discourse density scores, then the qualitative components of form for the resume profiles, before presenting how these features differ in resume profiles of different quality levels.

Disciplinary Discourse Density Fig. 1 exemplifies the calculation of disciplinary discourse density in resume profile entries. Identifying data such as company name, years of service, and very specific identifying information are redacted for anonymity, but otherwise, the textual data remain unedited to preserve capitalization, bullets, and indentation. We demonstrate the method of calculating disciplinary discourse density for these examples. (See Fillenwarth, McCall, and Berdanier [9] for more examples on coding in conventional resume contexts.)

The excerpt in Fig. 1(a), a "strong" resume profile, employing the equation for disciplinary discourse density (here applied to the entry, rather than the entire profile) would have a summed code score of 160 from a total of 46 codes, for a disciplinary discourse density score of 3.48. Similarly, the example in Fig. 1(b), also a strong resume, has a summed code score of 172 from 38 codes, for a disciplinary discourse score of 4.52. These examples, both from resume profiles deemed of high quality, show different approaches to style and rhetorical choices that can be applied in the same online resume profile tool.

Analysis of Qualitative and Quantitative Characteristics of Online Engineering Resume Profiles Structure within the resume profiles varied widely. As shown, the excerpt in Fig. 1(a) is oriented in the characteristic verb-centric language of conventional resumes, but formatted as a narrative paragraph, while that in Fig. 1(b) is in bullet format. To examine patterns of conventions

Criteria	Strong (3)	Moderate (2)	Weak (1)
Overall Appearance and Format *Goal:* Ensure résumé is well-structured and highlights skills, strengths, and experiences that are relevant to the employer	▪ Appropriate formatting ▪ Relevant information appears throughout résumé ▪ Section headings reflect content; content substantiates headings	▪ Some instances of inconsistent formatting ▪ Some relevant information throughout résumé ▪ Important information may not be clear or stand out to reader	▪ Résumé formatting is inconsistent throughout ▪ Information is not presented or articulated clearly ▪ Lack of relevant information throughout résumé ▪ Section headings do not accurately reflect content
Typos, Grammar, Spelling, and Style Errors	▪ Free of spelling, punctuation, and spacing errors ▪ Grammar is appropriate and consistent ▪ Consistent use of style	▪ Few and minor spelling, punctuation, or spacing errors ▪ Few instances of inconsistent style	▪ Résumé is difficult to understand due to numerous errors in spelling, punctuation, grammar, or spacing ▪ Many inconsistencies in style
Education *Goal:* To convey academic qualifications, training, or certifications	▪ Entries in reverse chronological order ▪ Academic qualifications (degrees, years, majors/minors, GPA) are clear ▪ Each institution includes name, location, dates of attendance ▪ Relevant training or certifications listed and complete	▪ Degree or institutional information is abbreviated and incomplete ▪ Academic qualifications are incomplete ▪ Some training or certifications are irrelevant	▪ Missing information related to institutions, degree, major, concentrations, or dates of attendance ▪ List institutions from which no degrees were earned ▪ Entries in an order other than reverse chronological order ▪ Statements or omissions seem misleading
Experience *Goal:* To highlight relevant experiences, skills, and accomplishments	▪ All positions are listed in reverse chronological order ▪ Positions include organization or company name, position title, location, and dates ▪ Highlighted experiences contain strong action words ▪ Results and impact are stated strongly and quantified when possible ▪ Use industry- and discipline- specific language correctly throughout the résumé	▪ Some positions are out of order ▪ Missing minor information related to organization or company ▪ Highlighted experiences are mostly relevant ▪ Results or impact could be stated more strongly or quantified ▪ Industry- and discipline-specific language is inconsistent at times or could be strengthened	▪ Order of experiences is illogical ▪ Missing substantial information about organization or company ▪ Highlighted experiences are vague, irrelevant, or contain weak language ▪ Results or impact is unaddressed or vague, not quantified when appropriate ▪ Little use of industry- or discipline-specific language throughout
Additional Sections *Goal:* Demonstrate other relevant experiences, skills, and accomplishments	▪ Listings relevant to targeted discipline ▪ Listings are concise	▪ Most listings are relevant ▪ Most listings are concise	▪ Listings are vague or irrelevant ▪ Listings are wordy

Fig. 2. Rubric for evaluating online resume profiles; adapted from Kain and Wardle [47].

TABLE III
DESCRIPTIVE STATISTICS FOR DISCIPLINARY DISCOURSE PATTERNS AND CONVENTIONS BETWEEN RESUME PROFILES OF DIFFERING QUALITY

Quality of Resume Profile	Mean Page Length	Mean Overall Summed Tier Score	Mean Disciplinary Discourse Density Score	Format: Using Primarily Bullets (% Split Between Bullets and Narrative)	Style: Using Phrases Starting with Verbs (% Split Between Verb-Centric and Other Phrasing)
Strong	$\bar{x}$ = 2.34 SD = 0.69	$\bar{x}$ = 516.77 SD = 248.19	$\bar{x}$ = 3.89 SD = 0.37	72.7% (18.1%)	81.8% (18.1%)
Moderate	$\bar{x}$ = 1.57 SD = 0.48	$\bar{x}$ = 211.23 SD = 120.76	$\bar{x}$ = 3.68 SD = 0.56	70.6% (14.7%)	70.6% (32.4%)
Weak	$\bar{x}$ = 1.01 SD = 0.51	$\bar{x}$ = 95.79 SD = 71.80	$\bar{x}$ = 3.48 SD = 0.52	53.6% (6.0%)	65.2% (31.8%)

in engineering resume profiles, which have not been captured before in the literature, we analyzed the qualitative components that accounted for the primary differences between web and conventional resumes per the categories in the rubric by which we categorized the resumes: length, format, and stylistic convention. The descriptive results are summarized in Table III.

Length of Engineering Resume Profiles Differs With Quality: We acknowledge that "length" is somewhat arbitrary for resume profiles, which do not have general length conventions. However, as a basis for comparing and characterizing typical conventions within resume profiles, we chose to employ the length of the resume as one of the features captured. To standardize the process, we saved each resume as a PDF by using the download feature of Indeed.com and did not change formatting or reduce spacing within the resulting files. Therefore, we could compare the *relative* length of the documents to get an understanding of how long these documents are when printed out by a prospective employer.

Default formatting does influence the length of the resume profiles, which arranges resumes such that the person's name, location, and hiring notes (e.g., willingness to move; international work visa status) fills approximately one-quarter of the first page, perhaps resulting in the spillover of resume credentials onto the following page. For this analysis, we rounded the page length up to the nearest quarter page from the downloaded PDF that could be printed. This decision was made because many of the weak resume profiles were half a page or three-quarters of a page long, a detail that could not be conveyed if page length were calculated by rounding up to the full page.

We sought to understand whether page length varied in statistically significant ways based on the quality of the resume profile. After testing for homogeneity of variance using Levene's test, a one-way ANOVA was conducted to compare the mean page length between the strong, moderate, and weak engineering resume profiles. There was a statistically significant difference at the $\alpha = 0.05$ level in the average page length [$F(2,86) = 38.55$, $p < 0.001$]. Post-hoc Tukey HSD analyses revealed that mean page length differed statistically significantly between the strong and moderate engineering resume profiles ($p < 0.001$, $d = 1.39$); between the moderate and weak resume profiles ($p < 0.001$, $d = 1.03$); and between the strong and weak resume profiles ($p < 0.001$, $d = 2.42$). This finding means that for the engineering resume profiles, the strong resumes were statistically significantly different between all three groups, with the strong profiles being longer than the moderates, and the moderates longer than the weak. The default formatting for resume profiles may influence this finding: The unlimited text format for the Indeed.com resume profile generator enables applicants to write without editing or conceptualizing the use of space as they might in a conventional resume. To this end, many of the resume profiles only partially filled their last page. We do not know whether the people feeding their information into the Indeed.com resume generator checked to see how their content is translated into printed text.

Disciplinary Discourse Density and Disciplinarity of Resumes Vary With Quality: After testing for

homogeneity of variance with Levene's test, a one-way ANOVA comparing the mean disciplinary discourse density between the strong, moderate, and weak engineering resume profiles revealed a statistically significant difference at the $\alpha = 0.05$ level in the average disciplinary discourse density score ($F(2,86) = 4.08$, $p = 0.02$). A post-hoc Tukey HSD analysis revealed statistically significantly differences between the strong and weak categories ($p = 0.02$, $d = 0.78$). This result means that strong engineering resume profiles had higher disciplinary discourse density (e.g., used disciplinary language that met, on average, higher tier categories from the AAES model) than the weak resumes.

As the variances in the mean overall tier scores between the groups were found to be heterogeneous using Levene's test, a Welch's ANOVA test was employed. Results showed that there were statistically significant differences at the $\alpha = 0.05$ level in the mean overall tier score between the groups ($F(2,56.48) = 5.35$, $p = 0.007$). A Games-Howell post-hoc analysis revealed that the mean overall tier score differed statistically significantly between the strong engineering resume profiles and the weak profiles ($p = 0.005$, $d = 0.88$). Although there were no statistically significant differences between the strong and moderate profiles or between the moderate and weak profiles, these findings show that the stronger resume profiles tended to have higher summed tier scores over the entirety of each resume.

Rhetorical Style and Format of Resume Content Vary With Quality: In determining format, we categorized each resume in terms of whether it was *primarily* structured using bullets. There were a few resumes that used numbered lists instead of bullets or wrote short phrases on separate lines but did not use actual bullets. Other people used symbols such as an asterisk or equal sign in place of bullets. All of these scenarios were counted as bullets because they were non-narrative in format (e.g., not a block paragraph describing a person's position). Some resumes were split such that there was not a true primary format. We described these as "split format" and we noted the percentage of split-formatted documents in parentheses in Table II. Trends indicate that only about half of the "weak" resume profiles were primarily in bullet format. "Moderate" resumes more often used bullet format, with nearly 70% subscribing primarily to bullet format, and nearly 73% of "strong" resume profiles primarily used bullets.

To analyze whether the use of conventions differed significantly with quality of resume, we performed a nonparametric chi-square test for independence at the $\alpha = 0.05$ level and determined that the differences between categories did not occur by chance, χ^2 (4, N = 89) = 0.05, $p < 0.001$. In sum, the quality of a resume is linked with the likelihood of using bullets, with moderate resume profiles most likely to employ bullet formatting.

The rhetorical style of sentences and phrases within the resume profiles also varied. Approximately 65% of weak resume profiles used sentences or phrases that started with verbs, with an additional 31% employing a split style comprised equally of phrases that start with verbs and phrases that started with other syntactic elements. Moderate resume profiles employed verb-centric phrases in approximately 70% of the sample, whereas ~82% of strong profiles employed a predominantly verb-centric approach. For both the moderate and strong profiles, a substantive number of the remaining resumes were in a split format.

To analyze whether the use of rhetorical styles differed significantly with the quality of resume, we performed a nonparametric chi-square test for independence at the $\alpha = 0.05$ level. We determined that the differences between categories did not occur by chance, χ^2 (4, N = 89) = 0.71, $p = 0.001$. Therefore, the quality of resume is statistically linked with the likelihood of using verb-centric phrases, with strong resumes usually employing verb-centric phrasing and weak resumes usually employing sentences to describe their roles.

Although the assumption may be that the resumes that did not employ bullets may be more likely to use nonverb-centric language, this trend did not hold. For example, Fig. 1(a) shows an excerpt where the writer does not use bullet formatting but does use the verb-centric style in his or her resume (in the form of grammatically incorrect sentences). These "mismatches" between format and style were common across quality designations. Of the weak resume profiles, 39% used bullets with split or nonverb phrases; narrative formats with verb-based or split phrases; or split formatting and rhetoric together. Similarly, 32% of moderate and 36% of strong resume profiles were mismatched. This mismatching may be a result of web-editing, haste, or—potentially—a sentiment that a resume profile isn't a "real" resume (though we did not investigate rationale). Indeed, many of the weak

resumes seemed incomplete, hosting only the educational background of a potential candidate.

Discussion and Conclusion

While our prior research with conventional engineering resumes indicated that disciplinary discourse density could be an effective way to quantify and teach rhetorical strategy in engineering resumes, this research sought to extend the method to resume profiles. We posited that this application would be effective because, although the activity systems of web and conventional resumes are similar, they are nonequivalent except for the junction between the subject and the community. Therefore, disciplinary discourse density represents the connection between the job seekers and the engineering disciplinary community, represented through the choice of words and phrases that mean something particular to that community. These do not change with the venue of a resume profile. Indeed, our statistical analysis found that there is a statistically significant difference between quality levels for both the overall tier scores and the mean disciplinary discourse scores for engineering resume profiles.

These findings emphasize the importance of disciplinary language in both resume research and pedagogy. With disciplinary discourse remaining stable across the subject–community junction for both traditional and resume profile activity systems, it is imperative for researchers and instructors to explore the role of lexical choices in resume quality. Although scholars like Diaz [33] have begun to move in this direction by discussing how choices regarding language may influence readers, there is room for more robust research and development of teaching practices regarding language use in resumes and resume profiles. Especially given the wide-ranging, continually changing array of print and digital contexts in which resume content may appear, it is essential for researchers, instructors, and writers to understand the role that language—as one of the only constants among resume writing activity systems—plays as applicants showcase their qualifications and demonstrate their fit within a disciplinary community. In terms of helping students analyze the rhetorical situations for which they are producing resumes, exercises that ask students to actively investigate the activity systems at play in the workplace and in the job application process, such as those described in Kain and Wardle [47] and Randazzo [1], may prove to be particularly useful pedagogical tools.

From an activity theory perspective, this research also highlights significant shifts in the nodes in the activity systems of conventional resumes and resume profiles. The division of labor for the resume profile's activity system represents a particularly notable role reversal. For conventional print resumes, the burden of argumentation is on the resume writer, who is responsible for choosing among relevant experiences and describing them in a way that will align with an employer's job ad and expectations. In the employment website environment, the resume writer is targeting a broad range of jobs and employers. Rather than carefully curating one's experiences for a particular audience, the resume writer instead presents as much information about him or herself as deemed relevant. The employer, then, is responsible for determining to what extent a candidate's resume reveals a capacity to succeed in a job position—whether relying on human readers or applicant tracking software.

This change in roles in the division of labor highlights particularly well the constitutive function of the genre; that is, the way that genres make possible (and, in this case, demand) certain activities. (See Bawarshi [59] for a fuller discussion.) By providing a much broader and less succinct overview of a candidate's qualifications, the genre of the resume profile necessarily shifts how the employer will approach the resume review process. As a result of this shift, researchers and educators must consider, and teach students how to consider, the differences in how recruiters will approach the task of reviewing applicants based on the way resumes are received. Although numerous past studies have researched employers' perceptions of resumes and job applicants [60]–[63], it becomes evident that additional research is needed to understand how employers process and assess resumes in this new activity system.

Implications for Teaching With the emerging differences in activity systems between web and conventional resumes, we as instructors must question whether some conventions of the resume from the era before resume profiles (e.g., the one-page standard for entry-level resumes and the amount of detail provided about experience) need to be reconsidered. From an activity theory perspective, the division of labor for resume profiles shifts from the applicant to a generic employer, such that the resume is no longer a tailored artifact for a specific company.

At the same time, though, in teaching resume profiles, we must also help students navigate this new activity system and the "more is better" position that platforms like Indeed.com may invite. With open-style text boxes and a lack of strict length limits, our results indicate that applicants used this space, but not necessarily in the most productive way. Although it is possible to create a resume profile that is both long (compared to conventional resume standards) and strong, we observed many resumes, particularly in the "moderate" category, that did not use the added length to their advantage. It will be important to help students learn how to utilize this additional space strategically. Teaching the importance of disciplinary discourse patterns for both conventional print resumes and resume profiles can help students develop a heightened rhetorical sensibility, enabling them to identify specific engineering audiences and to make language choices that best convey their specific engineering expertise to these audiences.

Another area of interest to researchers, instructors, and students involves resume organization and design. As noted above, many of the stylistic and formatting changes present in the resume profile venues result from the fill-in format of websites like Indeed.com. For researchers and teachers, this predetermined format is an important change to take into consideration and address with students. With the use of prebuilt forms, classroom discussions can shift away from the importance of including required elements like job title and dates of employment and toward how to work with a "description" box to clearly convey experiences, expertise, and achievements both verbally and visually.

Though resume design has long been a topic of study (e.g., [11], [17], [21], [22], [33]), the standardized format of resume profiles also notably eliminates opportunities for personalization and the display of design skills, therefore nullifying many of the rhetorical strategies that resume writers can access in conventional forms of the genre. This de-emphasis on design calls for extra attention to the importance of clear language that meets disciplinary expectations to stand out from competing candidates. In the classroom, instructors are compelled to prepare students for the range of resume writing activity systems—conventional and emergent—that they will encounter throughout their careers. Teaching "the resume" as a monolithic, print-based genre is no longer a viable approach.

Limitations and Opportunities for Future Work

As with any study, there are limitations to the reach of the study that motivate opportunities for future theoretical and empirical work relating to conventional resumes and resume profiles. The first limitation is the sample size. Although the sample sizes are appropriate for qualitative data analysis (indeed, $N = 89$ is a great deal larger than most qualitative corpuses), it would be interesting to extend to a much larger sample size of engineering resumes or to extend across resume profile platforms to generalize findings more extensively.

Because of the research questions that we asked, this study did not investigate whether there are differences in how web-based resumes operate in relation to a user's conventional print resume. In other words, we do not know whether users digitally copied and pasted information from a conventional resume into the web form (e.g., *transforming* their conventional resume into a resume profile) versus those who used the Indeed.com platform to *generate* a resume from scratch. Future studies could compare the same participants' conventional resume and resume profiles to highlight potential differences.

We also did not capture whether the participants designed their online resume profiles to optimize an ATS or the extent to which they felt such optimization was necessary. Similarly, we did not explore how the employers or other hiring decision-makers employ web-based platforms like Indeed.com. Therefore, an opportunity for future work is to explore the relationship between making job materials "compatible with both electronic and human requirements" [29, p. 18], including how employers actually read and interpret resume profiles, through actor-network theory (ANT), which is a theoretical and methodological approach to understanding the "social" not as a homogeneous entity, but as a "trail of *associations* between heterogeneous elements" [64, p. 5, emphasis in the original].

One of ANT's most discussed features is its insistence on the symmetry between human and nonhuman actors—actants—whose assemblage creates a network of relations [65]–[67]. A follow-up study could explore these research topics with engineering employers and hiring stakeholders. This might also be an opportunity to critically reflect on the role of ATSs as actants and gatekeepers within a disciplinary hiring process.

Concluding Thoughts In this study, we posited that conventional resumes and resume profiles belong to the same genre but are produced in different activity systems and different media, a fact that leads to similar yet divergent genre conventions. The findings from our study not only validate the significance of disciplinary discourse and use of disciplinary discourse density in considering quality within web engineering resumes but also characterize stylistic conventions that differ from those of conventional resumes. We showed that the use of disciplinary discourse and various format and style conventions differ statistically significantly with the quality of the resume profiles. We situate the qualitative differences between the two within the parallel, but distinct components of their respective activity systems. The differences between the two activity systems in which conventional and web-based engineering resumes are deployed can highlight areas for potential educational interventions for technical and professional communication instructors who teach resume writing practices to students for conventional and online venues.

As the web continues to grow as a dominant tool used in job searches and recruitment, it will be critical to continue studying the resume's evolution as a genre. Of interest will be exploring whether the conventional resume and resume profile eventually grow apart in audience, purpose, content, style, and design such that these types of resumes move from having a handful of divergent genre conventions to eventually becoming two entirely distinct genres. This continued careful attention to the rhetorical function of the resume and related job application documents—the same attention that we seek to cultivate in our students—is essential to preparing engineering students, as well as students from a range of professional and technical majors, for the job search and their future careers.

Fig. 2 shows the rubric used to evaluate online engineering resume profiles.

References

[1] C. Randazzo, "Where do they go? Students' sources of résumé advice, and implications for critically reimagining the résumé assignment." *Tech. Commun. Quart.,* vol. 45, no. 4, pp. 278–297, 2016.
[2] J. B. Killoran, "Self-published web résumés: Their purposes and their genre systems," *J. Bus. Tech. Commun.*, vol. 20, no. 4, pp. 425–459, 2006.
[3] C. M. Hanchey, "Creating effective résumés," *J. Comput. Sci. Coll.*, vol. 20, no. 4, pp. 126–126, 2005.
[4] T. Krause, "Preparing an online resume," *Bus. Commun. Quart.*, vol. 60, no. 1, pp. 159–162, 1997.
[5] E. Furtmueller, C. Wilderom, and R. M. Mueller, "Online résumés: Optimizing design to service recruiters," in *Proc. 18th Eur. Conf. Inf. Syst.*, vol. 53, pp. 1–12, 2010. [Online]. Available: http://aisel.aisnet.org/ecis2010/53
[6] J. Guillory and J. T. Hancock, "The effect of Linkedin on deception in resumes." *Cyberpsych. Beh. Soc. Netw.*, vol. 15, no. 3, pp. 135–140, 2012.
[7] R. H. Beatty, *The Resume Kit (5th ed.)*, Hoboken, NJ, USA: Wiley, 2003.
[8] M. Balachandran and R. B. Blair, "Résumé preferences of US businesses," *Bus. Educ. Dig.*, vol. 16, 2007.
[9] G. M. Fillenwarth, M. McCall, and C. G. P. Berdanier, "Quantification of engineering disciplinary discourse in resumes: A novel genre analysis with teaching implications," *IEEE Trans. Prof. Commun.*, vol. 61, no. 1, pp. 48–64, Mar. 2018.
[10] K. L. Hutchinson and D. S. Brefka, "Personnel administrators' preferences for résumé content: Ten years after," *Bus. Commun. Quart.*, vol. 60, no. 2, pp. 67–75, 1997.
[11] B. Culwell-Block and J. A. Sellers, "Résumé content and format—Do the authorities agree?" *Bus. Commun. Quart.*, vol. 57, no. 4, pp. 27–30, 1994.
[12] P. Stanley-Weigand, "Organizing the writing of your resume," *Bull. Assoc. Bus. Commun.*, vol. 54, no. 3, pp. 11–12, 1991.
[13] J. Harcourt and A. C. Krizan, "A comparison of resume content preferences of fortune 500 personnel administrators and business communication instructors," *J. Bus. Commun.*, vol. 26, no. 2, pp. 177–190, 1989.
[14] C. P. Bird and D. D. Puglisi, "Method of resume reproduction and evaluations of employment suitability," *J. Bus. Commun.*, vol. 23, no. 3, pp. 31–39, 1986.
[15] K. L. Hutchinson, "Personnel administrators' preferences for resume content: A survey and review of empirically based conclusions," *J. Bus. Commun.*, vol. 21, pp. 5–14, 1984.
[16] J. M. Penrose, "A discrepancy analysis of the job-getting process and a study of resume techniques," *Int. J. Bus. Commun.*, vol. 21, no. 3, pp. 5–15, 1984.
[17] J. M. Penrose, "Does order make a difference in resumes?" *Int. J. Bus. Commun.*, vol. 10, no. 3, pp. 15–19, 1973.
[18] C. Martin-Lacroux and A. Lacroux, "Do employees forgive applicants' bad spelling in resumes," *Bus. Prof. Commun. Quart.*, vol. 80, no. 3, pp. 321–335, 2017.
[19] K. Tillotson and D. Osborn, "Effect of a resume-writing workshop on resume-writing skills," *J. Employment Couns.*, vol. 49, pp. 110–117, 2012.
[20] E. W. Wright, T. A. Domagalski, and R. Collins, "Improving employee selection with a revised resume format," *Bus. Commun. Quart.*, vol. 74, no. 3, pp. 272–286, 2011.

[21] N. M. Schullery, L. Ickes, and S. E. Schullery, "Employer preferences for resumes and cover letters," *Bus. Commun. Quart.*, vol. 72, no. 2, pp. 163–176, 2009.
[22] E. Blackburn-Brockman and K. Belanger, "One page or two?: A national study of CPA recruiters' preferences for resume length," *J. Bus. Commun.,* vol. 38, no. 1, pp. 29–57, 2001.
[23] M. Waung, R. W. Hymes, and J. E. Beatty, "The effects of video and paper resumes on assessments of personality applied social skills, mental capability, and resume outcomes," *Basic Appl.. Soc. Psychol.*, vol. 36, pp. 238–251, 2014.
[24] A. Anders, "Flipping the composing process: Collaborative drafting and resume writing," *Bus. Prof. Commun. Quart.,* vol. 79, no. 1, pp. 102–118, 2016.
[25] C. Randazzo, "Positioning resume and cover letters as reflective-reflexive process," *Bus. Commun. Quart.*, vol. 75, no. 4, pp. 377–391, 2012.
[26] R. Darolia, C. Koedel, P. Martorell, K. Wilson, and F. Perez-Arce, "Race and gender effects on employer interest in job applicants: New evidence from a resume field experiment," *Appl. Econ. Lett.*, vol. 23, no. 12, pp. 853–856, 2016.
[27] D. S. Furbish, "Exploring social justice via resumes," *J. Employment Couns.*, vol. 52, pp. 36–47, 2015.
[28] B. D. Davis and C. Muir, "Resume writing and the minority student," *Bus. Commun. Quart.,* vol. 66, no. 3, pp. 39–50, 2003.
[29] W. H. Baker, K. DeTienne, and K. L. Smart, "How fortune 500 companies are using electronic resume management systems," *Bus. Commun. Quart.*, vol. 61, no. 3, pp. 8–19, 1998.
[30] C. Roever, "Preparing a scannable resume," *Bus. Commun. Quart.*, vol. 60, no. 1, pp. 156–159, 1997.
[31] T. Krause, "Preparing an online resume," *Bus. Commun. Quart.*, vol. 60, no. 1, pp. 159–161, 1997.
[32] J. Kalmbach, "Hypermediating the resume," in *Innovative Approaches to Teaching Technical Communication*, T. Bridgeford, K. A. Kitalong, and R. Selfe, Eds. Logan, UT, USA: Utah State Univ. Press, 2004, pp. 135–147.
[33] C. S. Diaz, "Updating best practices: Applying on-screen reading strategies to resume writing," *Bus. Commun. Quart.*, vol. 76, no. 4, pp. 427–445, 2013.
[34] C. Randazzo, "A framework for resume decisions: Comparing applicants' and employers' reasons," *Bus. Prof. Commun. Quart.*, vol. 83, no. 4, pp. 409–433, 2020, [Online]. Available: https://doi.org/10.1177/2329490620963133.
[35] C. Morris. (2020). Strategies to beat the applicant tracking system and land the interview. [Online]. Available: https://www.indeed.com/career-advice/finding-a-job/expert-advice-applicant-tracking-system#:~:text=Applicant%20tracking%20systems%20save%20employers,information%20it%20includes%20to%20decide
[36] D. H. Charney, J. R. Rayman, and L. Ferreira-Buckley, "How writing quality influences readers' judgements of resumes in business and engineering," *J. Bus. Tech. Commun.*, vol. 6, no. 1, pp. 38–74, 1992.
[37] R. Valdez-Almada, O. M. Rodriguez-Elias, C. E. Rose-Gomez, M. De, J. Velazquez-Mendoza, and S. Gonzalez-Lopez. "Natural language processing and text mining to identify knowledge profiles for software engineering positions: Generating knowledge profiles from resumes," in *Proc. 5th Int. Conf. Softw. Eng. Res. Innov.*, 2017, pp. 97–106.
[38] S. R. Elliott, "Determination of key engineering competency attributes by evaluating resumes using machine learning," Ph.D. dissertation, Dept. Eng. Manage. Syst. Eng., The George Washington Univ., Washington, DC, USA, 2021.
[39] L. S.Vygotsky, *Thought and Language.* A. Kozulin, Ed., Trans. Cambridge, MA, USA: MIT Press, 1986 (Original work published 1934).
[40] L. S. Vygotsky, *Mind in Society*, Cambridge, MA, USA: Harvard Univ. Press, 1978.
[41] A. N. Leont'ev, *Activity, Consciousness, and Personality*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1978.
[42] A.N. Leont'ev, *Problems of the Development of Mind.* Moscow, Russia: Progress, 1981.
[43] Y. Engeström, *Learning By Expanding: An Activity Theoretical Approach to Developmental Research*. Helsinki, Finland: O.-K. Oy, 1987.
[44] M. Cole and Y. Engeström, "A cultural-historical approach to distributed cognition," in *Distributed Cognitions: Psychological and Educational Considerations,* G. Salomon, Ed. Cambridge, UK: Cambridge Univ. Press, 1993, pp. 1–46.
[45] M. Cole, *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA, USA: Harvard Univ. Press, 1996.
[46] D. R. Russell, "Rethinking genre in school and society: An activity theory analysis," *Write Commun.*, vol. 14, no. 4, pp. 504–554, 1997.
[47] D. Kain and E. Wardle, "Building context: Using activity theory to teach about genre in multi-major professional communication courses," *Tech. Commun. Quart.*, vol. 14, no. 2, pp. 113–139, 2005.
[48] Y. Engeström, "Activity theory and individual social transformation," in *Perspectives on Activity Theory,* Y. Engeström, R. Miettinen, and R-L. Punamaki. Eds. Cambridge, UK: Cambridge Univ. Press, 1999, pp. 19–38.
[49] V. Kaptelinin, "The object of activity: Making sense of the sense-maker," *Mind, Culture, Activity,* vol. 12, pp. 4–18, 2005.
[50] C. Spinuzzi, "Losing by expanding: Corralling the runaway object," *J. Bus. Tech. Commun.*, vol. 25, no. 4, pp. 449–486, 2011.
[51] B. A. Nardi, "Activity theory and human-computer interaction," in *Context and Consciousness: Activity Theory and Human-Computer Interaction*, B. A. Nardi, Ed. Cambridge, MA, USA: MIT Press, 1996, pp. 7–26.
[52] A. S. Bawarshi and M. J. Reiff, *Genre: An Introduction to History, Theory, Research, and Pedagogy*. West Lafayette, IN, USA: Parlor Press, 2010.
[53] Y. Engeström, *From Teams to Knots: Studies of Collaboration and Learning at Work.* Cambridge, UK: Cambridge Univ. Press, 2008.

[54] Y. Engeström, R. Engeström, and T. Vähääho, "When the center does nothold: The importance of knotworking," in *Activity Theory and Social Practice*, S. Chaiklin, M. Hedegaard, and U. J. Jensen, Eds. Aarhus, Denmark: Aarhus Univ. Press, 1999, pp. 345–374.
[55] C. G. P. Berdanier, M. McCall, and G. Mike, "Resumes in the development of undergraduate engineering identity: A genre analysis with teaching implications," in *Proc. IEEE Int. Commun. Conf.*, 2016, pp. 1–9.
[56] M. McCall, G. M. Fillenwarth, and C. G. P. Berdanier, "Quantification of disciplinary discourse: An approach to teaching engineering resume writing," in *Diverse Approaches to Teaching, Learning, and Writing Across the Curriculum: IWAC At 25*, L. E. Bartlett, S. L. Tarabochia, A. R. Olinger, and M. J. Marshall, Eds. Fort Collins, CO, USA: WAC Clearinghouse, 2020, pp. 113–134.
[57] American Association of Engineering Societies Engineering. Competency model. (2016). [Online]. Available: https://www.careeronestop.org/competencymodel/competency-models/engineering.aspx
[58] *Engineering Professional Development Career Resource Manual*. Iowa City, IA, USA: University of Iowa College of Engineering. 2015. [Online]. Available: http://www.pageturnpro.com/University-of-Iowa-College-of-Engineering/67540-Career-Resources/index.html#10
[59] A. S. Bawarshi, "The genre function," *Coll. English*, vol. 62, pp. 335–360, 2000.
[60] D. H. Charney and J. R. Rayman, "The role of writing quality in effective student resumes," *J. Bus. Tech. Commun.,* vol. 3, no. 1, pp. 36–53, 1989.
[61] M. S. Cole, R. S. Rubin, H. S. Feild, and W. F. Giles, "Recruiters' perceptions and use of applicant resume information: Screening the recent graduate," *Appl. Psychol.*, vol. 56, no. 2, pp. 319–343, 2007.
[62] E. Derous, "Investigating personnel selection from a counseling perspective: Do applicants' and recruiters' perceptions correspond?" *J. Employment Couns.*, vol. 44, no. 2, pp. 60–72, 2007.
[63] J. K. Arnulf, L. Tegner, and Ø. Larssen, "Impression making by resume layout: Its impact on the probability of being shortlisted," *Eur. J. Work Organiz. Psychol.,* vol. 19, pp. 221–230, 2010.
[64] B. Latour, *Reassembling the Social: An introduction to Actor-Network-Theory.* New York, USA: Oxford Univ. Press, 2005.
[65] C. Spinuzzi, *Network: Theorizing knowledge Work in Telecommunications*. Cambridge, UK: Cambridge Univ. Press, 2008.
[66] B. Latour, *Aramis Or the Love of Technology*, C. Porter, Trans. Cambridge, MA, USA: Harvard Univ. Press, 1996.
[67] B. Latour, "On actor-network theory: A few clarifications," *Soziale Welt,* vol. 47*,* pp. 369–381, 1996.

Catherine G. P. Berdanier is an Assistant Professor of Mechanical Engineering at the Pennsylvania State University, State College, PA, USA. Her research interests include graduate-level engineering education, attrition, persistence, socialization, and engineering writing and communication.

Mary McCall is an Assistant Professor of English at North Dakota State University, Fargo, ND, USA. Her research interests include professional and technical writing, writing across the curriculum, and women's, gender, and sexuality studies.

Gracemarie Mike Fillenwarth is an Assistant Professor of Writing Arts at Rowan University, Glassboro, NJ, USA. Her research and teaching interests include technical and professional communication, writing across the curriculum, administration, and pedagogy, in both historic and present-day contexts.

Research Article

Information Design for Small Screens: Toward Smart Glass Use in Guidance for Industrial Maintenance

—Hanna Heinonen, Sanni Siltanen, and Petri Ahola

***Abstract*—Background:** *Smart glasses and other extended reality (XR) solutions provide new ways of utilizing technical documentation with hands-busy tasks in the field. Scaling up the use of XR solutions in industry has been difficult due to the manual authoring of content for each device and task. Therefore, authoring solutions and information design methods need to be developed to scale content automatically to different devices and applications.* **Literature review:** *Related work includes smart glasses and industrial maintenance work, categorization based on users' skill levels, and standardized guidelines in information design.* **Research questions:** *1. How should information content be designed and created to support use in smart glasses and other small-screen devices in addition to existing delivery channels? 2. How can the same information content be utilized to deliver relevant content to users based on their skill levels? 3. Are the users of technical instructions ready to accept smart glasses and XR as a delivery channel?* **Methodology:** *We describe a study that focused on designing maintenance instructions for small screens. The information was authored in DITA XML format, and a smart glass application was used in user tests to evaluate the delivery and usability of the information. We used thinking aloud and participant observation as well as questionnaires to collect data.* **Results and discussion:** *The chosen information design methods successfully compressed technical information, and automatic filtering of content supported different use cases. Participants were enthusiastic about the use of smart glasses, and the instructions helped in performing tasks.* **Conclusions:** *Information designed with the user-centered approach of minimalism works best with instructions on small screens, and filtering information using DITA XML elements is an efficient way to scale information for different user needs.*

Index Terms*—Darwin Information Typing Architecture (DITA), industrial maintenance, smart glasses, structured authoring, extended reality (XR).*

Traditionally, maintenance instructions have been delivered on paper copies, but paper delivery has proven problematic for many reasons. For example, an outdated copy of instructions might be accidentally used, and there are costs involved with printing. The trend has been, therefore, to move the delivery of information to electronic print formats—i.e., PDFs and more recently online portals. However, as industrial maintenance is often characterized by tasks that require the use of personal protective equipment while keeping the technician's hands busy holding equipment parts and tools, it is often difficult to access instructions while performing a task. Therefore, to solve this issue, the focus has turned beyond PDFs and online portals to new delivery methods, especially to extended reality (XR) solutions.

Manuscript received December 20, 2020; revised May 21, 2021; accepted May 22, 2021. Date of publication October 14, 2021; date of current version November 24, 2021. *(Corresponding author: Hanna Heinonen.)*
The authors are with the KONE Corporation, 05801 Hyvinkää, Finland (email: hanna.heinonen@kone.com; sanni.siltanen@kone.com; petri.ahola@kone.com).

IEEE 10.1109/TPC.2021.3110616

Extended reality is an umbrella term for technologies combining virtual and real elements. There are several levels of immersivity among XR applications, varying from total digital virtual reality (VR) environments, through mixed reality (MR), to augmented reality (AR), where the virtual elements are added on top of the user's view of reality [1]. In industry, VR is used, for example, in training and design reviews [2], and MR is used for many visualization purposes [3]. AR, especially when applied with smart glasses, is used in industrial settings for workflow guidance, reporting, and remote assistance [4], [5]. Solutions where information is displayed on smart glasses are sometimes also referred to as informed reality or assisted reality.

In industrial maintenance, it is not feasible in terms of resources to tailor instructions manually for different outputs and devices. Therefore, it is essential that the technical instructions displayed on XR devices be automatically retrieved from a company's repository. As demonstrated by Siltanen and Heinonen, Darwin Information Typing Architecture (DITA) XML is a good candidate for the creation of content for XR applications [6]. DITA is the technical communication standard for

Practitioner Takeaway

- Instructions displayed on smart glasses help users perform tasks in industrial maintenance, and attitudes towards using smart glasses are positive.
- Well-designed technical information can be automatically used and delivered for several use cases and in several end devices.
- Information designed with the user-centered approach of minimalism works best with instructions on small screens, and filtering information using DITA XML elements is an efficient way to scale information for different user needs.

Fig. 1. Test participants performing maintenance tasks during user tests. The personal protective equipment (hardhat, safety shoes, and cut-resistant gloves) is the same that is required for field operations.

structured writing, and it defines three main topic types: concept (what something is like or how it works), task (how something is done), and reference (information users might need when performing a task) [7]. A repository is used to store the topics, which can then be reused across publications and in different delivery channels.

However, even though we are nearing the era when XR solutions could be utilized on a wider scale in industrial maintenance, very little has been done to create any guidelines for the use of the new media and technologies in the field that specializes in designing and creating technical instructions and guidance, technical communication [8]. Even though DITA offers the technical capability to produce content for XR solutions, the focus needs to be put on designing the content so that it scales according to the media and users' needs. It is also of utmost importance that these concepts are developed and tested in real industry settings, taking into account the actual work environment and the challenges that arise from it. For example, the mandatory use of personal protective equipment (see Fig. 1) in many industrial roles has a substantial effect on how different devices and solutions can be utilized in the field [9]. For example, if safety glasses are required, AR glasses need to be compatible with them to meet the safety standards.

KONE is a global leader in the elevator and escalator industry [10]. It operates in more than 60 countries with approximately 30,000 field employees with varied levels of expertise. As new products are introduced often, even experienced field workers need detailed instructions for them. KONE publishes hundreds of new or revised instructions each year, and, therefore, single sourcing and reuse between information products and different outputs is essential. KONE has also been actively looking into novel ways of delivering technical information to field workers.

The aim of this research is to better understand how technical information content should be designed and created to support use in XR and other small-screen devices in addition to the existing delivery channels. Many industrial companies already produce instructions in DITA XML format that can be transformed into PDF or HTML formats or published to a web service for dynamic delivery to applications. However, the same content needs to be adapted to fit better in small display devices and to automatically filter content to support XR use cases. This study is a step toward creating guidelines for the design of content for omnichannel delivery, including smart glasses.

Next, we review the related work and present our research questions. Then, we explain the research methodology and the test arrangements. Finally, we describe the results and discuss future directions, and end the article with conclusions.

Literature Review

This exploratory study brings together research related to smart glasses and industrial maintenance work, categorization of users based on their skills levels, and standardized guidelines in information design. Smart glass use is a prominent area of research in industrial maintenance, and combined with information design, it offers a new way of delivering maintenance instructions to field workers. When delivering instructions to the limited space available on small screens, it is very important that the information be readable and fits the screen. Therefore, we also wanted to look into categorization of users' skills levels as a way of delivering the correct amount of information.

Smart Glasses and Hands-Busy Work in Industrial Maintenance Modern maintenance work is no longer only about using physical tools, such as screwdrivers, and getting measurements and values on physical gauges. Increasingly, technicians need to interact with digital information and information systems. Equipment information has moved to Internet of Things clouds, and technicians need to check parameters, equipment performance indicators, and measurement readings from information systems. Instructions are no longer only paper manuals but are stored in electronic format and accessed with mobile devices.

In many cases, industrial maintenance tasks require hands-busy actions and, in some cases, both hands are required most of the time [11]. Considering that the technician also needs to wear gloves that can be dirty and greasy, it is clear that a mobile phone with a touch screen user interface is not the optimal device to access information systems in these situations. For these use cases, smart glasses provide a potential solution. User interaction challenges and solutions in maintenance are very similar to some healthcare situations where both hands are needed, the hands might be dirty, and special attention needs to be paid to maintaining a sterile field. In healthcare, the use of smart glasses has also been studied to overcome these challenges [12].

Klinker et al. presented more than 20 use cases for smart glasses in maintenance [13]. One of the prominent use cases is providing workflow information for maintenance technicians. Industrial companies have experimented with smart glasses and, in several use cases, maintenance instructions are displayed using smart glasses [4], [5]. Studies show that users are able to perform faster and reduce error rates with such systems [14]–[16]. In addition, users feel more competent and satisfied [17], [18]. Adequate instructions provided with smart glasses instantly enhance the users' skills. Thus, technicians with different competence levels are able to complete maintenance tasks successfully.

Smart glass use has been often studied and tested in single-purpose use, and content has been tailor-made for each use case, a solution that is costly. Since smart glasses are still quite costly as well, especially those that meet industrial standards, the adaptation of smart glasses to industrial maintenance has been low despite the obvious benefits. However, if existing technical instructions already used for other formats can be reused for smart glasses and if the delivery is automated, the number of use cases increases drastically. Furthermore, when the work processes and practices are designed to support multipurpose use, the return on investment increases considerably. For example, smart glasses can be used to get remote assistance, read technical information, and utilize AR guidance (see, for example, SightCall [19]).

Expert–Novice People working in the field have different skill levels, varying from trainees to experts with lifelong experience. Experienced maintenance technicians know how to perform highly complex maintenance procedures [20]. In many cases, the knowledge they possess is tacit or tribal in nature. At the other end of the continuum,

novice users are anxious about making mistakes and want to get started quickly with the tasks at hand [11]. Advanced users may know and remember the task sequence in general and need only to check some details. They prefer to have detailed guidance only when performing tasks that occur rarely or that are new to them [21]. Novice users, on the other hand, need more generic guidance to be able to complete the task. In short, different types of users with varying skill levels have different requirements for technical information [22].

The classification of users into categories based on their skill levels is an established concept in technical communication. Hackos, for example, classifies users into expert performers, competent performers, advanced beginners, and novices [23]. These user profiles are used when creating and delivering content to meet the needs and demands of a specific target group. A user can be an expert for some tasks and novice for other tasks. Thus, the level of expertise and the need for guidance depend on the task at hand. Funk et al. use adaptive assistance for field workers with three different expertise levels–novice, advanced, and expert–to give each group a different level of guidance [24]. A common practice with DITA XML is that the filtering is done with conditions (audience = expert/novice), which requires that each sentence or segment in the DITA XML source be defined for expert-level users, novice-level users, or both.

Traditionally, many maintenance tasks have been standardized. For example, a certain set of preventive maintenance actions is performed at a set interval, and by working with the same equipment for years, the technician learns the tasks. However, the ongoing fourth industrial revolution is also changing the way maintenance work is performed. The Internet of Things, connectivity, and digitalization are transforming even the more traditional heavy industries, giving rise to concepts such as remote and condition-based maintenance. In these settings, the tasks vary from one piece of equipment and day to another, the maintenance technicians have a unique set of tasks to be performed at each maintenance visit, and it is no longer possible even for the experienced maintenance technician to know the tasks related to a certain preventive maintenance visit, for example. Therefore, there is an increased need for guidance even for the experts.

Standardized Guidelines in Information Design

Company style guides are created to provide consistency within each technical instruction and a larger set. In addition to setting standards for writing and formatting, style guides enforce guidelines for information design. The foundations for company guidelines lay in technical communication literature, where, for example, it is recommended that the number of steps in procedures be restricted to a maximum of nine [25], or that information is chunked into meaningful pieces to help readers make lengthy text more manageable to reduce the cognitive load [26]–[29]. These guidelines are based on an established theory in psychology: the magical number seven, plus or minus two. According to this theory, most people are capable of storing between five to nine items in their short-term memory [30]. In addition to technical communication, this theory has been applied to several aspects of daily life, for example, the length of telephone numbers, or recommendations for the maximum numbers of points in oral presentations [31].

Even though these guidelines are widely used in information design, it has also been argued that users of technical instructions do not have to memorize the steps of a procedure, but they typically read the steps one by one as they are performing the related tasks [25]. But with hands-busy maintenance procedures, it is not always possible to flip the pages of a paper manual or scroll down the screen to proceed one step at a time, and the users are forced to memorize a certain number of steps or chunks. Therefore, it has been safer to refrain from writing long procedures and keep them as short as possible or meaningfully chunked, even if that ends up fragmenting the overall task.

Research Questions

As the literature is lacking in guidelines for small-screen information design, we wanted to validate the type of information design principles that should be used for small screens where the instructions are available while the user is performing tasks. Therefore, we designed our study to explore whether the established guidelines in technical communication apply when designing information specifically for small screens. We also experimented with the standard semantic structure of DITA XML to see whether it can be used to filter content based on the user's expertise level. Furthermore, since the use of smart glasses to

TABLE I
SELECTED TASKS AND RELATED DEVICES

Task Name	Device Used	Task Description
Install motion sensor cable	Connectivity device	Performed on the elevator car roof. The user routes the motion sensor cable through the roof into the elevator car.
Power-up Connection	Connectivity device	Performed on the elevator car roof. The user connects the motion sensor cable to a connectivity device, switches on the device, and checks that it turns on.
Replace MediaPlayer	Content streaming device	Performed in a residential building lobby. The user replaces a broken content streaming device behind a TV.

TABLE II
SELECTED INFORMATION DESIGN METHODS

Method	Description
Conventional	Based on original instructions available for the tasks. No redesign of content, but due to limitations of the test device used (the screen size and features of the application), some minor, mostly layout modifications were needed for both text and graphics. (Fig. 4, left)
Visual manual	Based on the visual manual theory of Gattullo et al. [32]. The aim is to replace text with graphics, symbols, and icons. (Fig. 4, center)
Minimalist	Based on the minimalism heuristics of van der Meij and Carroll [33]. With minimalism, be aware of what to include, but also what not include, using a user-centered approach. In practice, provide only the information that users need to perform a specific task, and omit unnecessary details. Error prevention information is also included to aid the user. (Fig. 4, right)

deliver technical instructions is new, we also wanted to study users' attitudes toward this new delivery channel.

Therefore, our research was motivated by three key research questions.

RQ1. How should information content be designed and created to support use in smart glasses and other small-screen devices in addition to existing delivery channels?

RQ2. How can the same information content be utilized to deliver relevant content to users based on their skill levels?

RQ3. Are the users of technical instructions ready to accept smart glasses and XR as a delivery channel?

RESEARCH METHODOLOGY

Test Material Design Authentic KONE elevator maintenance and installation tasks were utilized in the study. We selected tasks for which the test participants would have to follow the instructions and would not know or guess the steps involved, but which would be simple enough to be completed in the chosen test setup and within a set timeframe. Three different sets of task instructions were created for each of the selected tasks. The selected tasks are presented in Table I and the selected information design methods in Table II.

The conventional version was chosen for this study as the baseline. Earlier research had indicated that traditional technical content is too long for small-screen delivery [6], so we used this as the reference version for the study.

Second, a visual manual version was chosen because the underlying theory considers the whole documentation production process, including the conversion of existing documents to AR (see Fig. 2). It is designed both to convert existing paper-oriented instructions to augmented reality instructions and to write new AR instructions from scratch [32], [34], [35]. The theory promotes replacing existing text with graphics, symbols, and

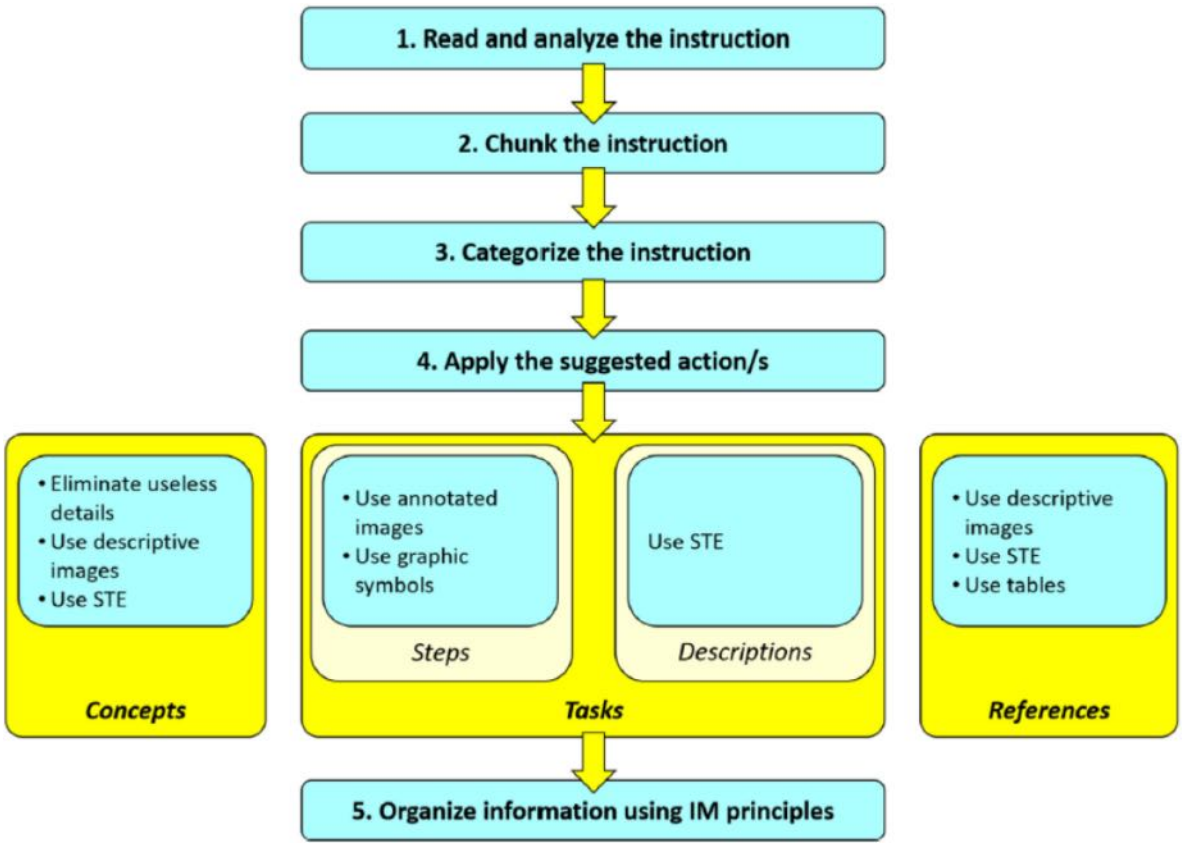


Fig. 2. Main conversion steps of visual manual theory (STE = Simplified Technical English, IM = Information Mapping) [32].

icons [32]. For many companies, legacy documentation must be utilized in one way or another to avoid losing their investment in it, even to the extent that the legacy and any new designs must be synchronized, since both are required by the company's field operations. As the amount of text in this version is limited, it is also cost-effective in terms of translation [34]. According to Gattullo et al., even though modern research acknowledges that more visual instructions are aligned with augmented reality (AR) and Industry 4.0 purposes, there is a lack of specific guidelines to convert existing instructions to visual manuals [32]. Most related research—for example, Knopfle et al. [36], Stock et al. [37], Engelke [38], Gimeno et al. [39], and Erkoyuncu et al. [40]—primarily discusses creating new content. Therefore, existing documentation does not reach technicians.

The third approach, using the minimalism heuristics of van der Meij and Carroll [33], was chosen because it is user-centric. It does not rely on conversion theory as such, but design principles for any technical communication content. In our test material design, the minimalism heuristics were used as a more researched counterpart to the visual manual theory. The minimalism heuristics work well on a small screen because unnecessary descriptive and transitional text is omitted.

The minimalism heuristics stress that one does not write instructions describing the full system, but includes only the steps that a user needs to perform his or her tasks. The user's goals and the tasks themselves must be known before writing the instructions. Also, error prevention information (notes, warnings) is added to save the user's time.

All the test content was authored in English in DITA XML format. DITA XML is also the standard suggested by Gattullo et al. for their visual manual theory [32]. Fig. 3 shows one of the original tasks, rendered in PDF format.

The original task explains why the cable has to be secured with cable ties, but for the target-audience, the trained maintenance technician, there is no need to include this information. As the conventional version was based on the wording of the original task, the explanation was left in that

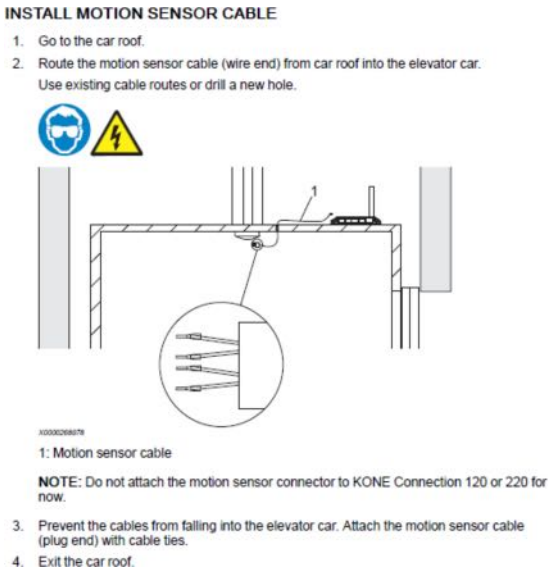


Fig. 3. Original version of the task "Install motion sensor cable."

version. For the visual manual version, we created an action symbol to indicate securing the cable and replaced the textual instruction with the symbol. For the minimalist version, we removed most of the descriptive text because, based on a target-audience analysis, we concluded that the target audience does not need it. Fig. 4 illustrates the results of step 3 of the original task with different methods. Test material design is described in detail in a master's thesis by Petri Ahola [41].

Test Application To test the concepts, a simple test application was developed for RealWear HMT-1 (see Fig. 5). We selected Realwear HMT-1 as the end device because it has a quite reliable voice user interface that facilitates hands-busy types of maintenance tasks. Furthermore, HMT-1 is compatible with a hardhat and a helmet, and it is a good representative of smart glasses suitable for industrial use.

In the test application, we targeted two issues:

- Focused information for experts, showing only certain parts of the existing content to the users—checklist type of use
- Information flow and the way the information is displayed on the screen

We wanted to show checklist type of information to people who are already familiar with basic maintenance tasks—that is, only the information that the expert-lever user needs. In this research, we use the term *checklist* to indicate a simple list of

Fig. 4. Three versions of the same step in the smart glass application: conventional (left), visual manual (middle), and minimalist (right).

items that you have to inspect or check, or simple steps that you have to perform, without lengthy explanations or descriptions of the steps. The intention was to tackle the problem that if we show all the information that a novice needs, the experts have problems finding the information that they need as the details are buried in the wealth of information. With our checklist, experts would have a reminder of what to check or do, but they could use their own expertise to perform the actual tasks.

We achieved the checklist by defining certain DITA XML elements that would be displayed; the rest would be ignored by the application (see Fig. 6). As DITA XML is based on semantic tagging of content [42], we could rely on the fact that certain elements would contain the information that we wanted to show. The approach that we utilized enables the reuse of information, and no hand-tailored expert files are needed. In other words, the same XML file can be used to resolve the full content, including the novice-level information, in another application or output. Traditionally, the expert-novice distinction is achieved with the use of conditions and conditional processing, but we wanted to

Fig. 5. RealWear HMT-1 with a hardhat.

```
<title>Install motion sensor cable</title>
<taskbody>
<prereq>You must have a new replacement cable before starting this task.</prereq>
<steps>
<step><cmd>Go to the elevator car roof.</cmd>
<info>Follow safe procedures for going to the car roof. Make sure you have familiarized yourself with all the needed safety instrutions.
<image href="W13.svg"></image></info>
</step>
<step><cmd>Route the motion sensor cable (wire end) from car roof into the elevator car.<image href="image1.png" placement="inline"></image></cmd>
<info>Pay attention to the ends of the cable and route it in the correct way.</info>
<stepresult>The cable is now routed properly.</stepresult>
</step>
<step><cmd>Prevent the cables from falling into the elevator car.</cmd>
<info><note>Watch out for other people working with the equipment. Always follow safe procedures when working on the elevator car roof.</note></info>
</step>
<step><cmd>Attach the motion sensor cable with cable ties.<image href="image2.png" placement="inline"></image></cmd>
<info>Make sure that the cable tie is firmly attached. Cut off extra lenght from the cable tie.</info></step>
<step><cmd>Exit the car roof.</cmd>
<info>Follow safe procedures for exiting the car roof. Make sure you have familiarized yourself with all the needed safety instrutions.
<image href="W13.svg"></image></info>
</step>
<step><cmd>Finalize maintenance visit.</cmd>
<info>Remove all the tools and waste material from the work site. Dispose of any waste according to local regulations.
<image href="image3.png" placement="break"></image></info></step>
</steps>
<result>You have now installed the motion sensor cable.</result>
</taskbody>
```

Fig. 6. Example of XML elements displayed and ignored by the application. Displayed elements are highlighted in the figure; all other fields are omitted.

explore ways to utilize the semantic structure of DITA XML and avoid adding conditioning to multiple parts of each topic. This approach saves time in the authoring phase as the conditions would always have to be manually assigned for each element.

Interface design for the test application was not in the focus of our study. However, some simple interactions were developed for the test application. The application was designed to show one step at a time, progressing with “next” and “back” voice commands. On each screen, a progress bar was shown that indicated the progression of the total maintenance task. The aim was to avoid or minimize the need for scrolling that is a usability problem on small screens, especially with voice commands [6]. In case the content did not fit on the screen, “page up” and “page down” commands could be used to access it. For an example of a UI screen, see Fig. 7.

We produced the XML files with XML authoring tools, and they had exactly the same structure as

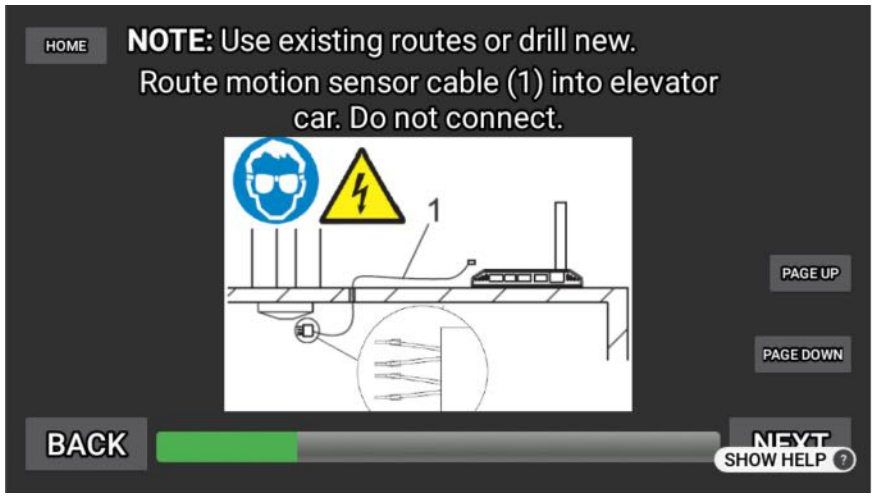


Fig. 7. Application UI showing one step, voice-activated action buttons, and the progress bar.

any XML files utilized in the company's technical instructions. For this study, the files were saved locally in the HMT-1. This allowed for the quick iteration of the different test files in the development phase as no publishing to the web service was needed. The tasks were accessible from a main menu in the application. The menu was created dynamically from an XML file containing references to the XML task files.

Participants A total of 21 test participants from the elevator company participated in the user tests and evaluated the different instructions (18 male, 3 female). Most participants fell into the 30–39 ($N = 8$) or 50–59 ($N = 7$) age groups, and the rest were from 40–49 ($N = 4$), 20–29 ($N = 1$), or > 59 ($N = 1$) age groups. All the participants were native Finnish speakers but had at least a working knowledge of English.

Six of the participants had worked in a field position or were very experienced, two were familiar with the field environment, and six were somewhat familiar with the field environment. The rest of the participants had very limited experience of the field environment ($N = 2$) or no experience ($N = 6$). However, the participants that had limited or no personal experience of the field environment were working with field documentation, maintenance development, or related roles, and all but three can be considered expert-level users of maintenance instructions.

Most of the participants had experience in industrial maintenance ($N = 10$) or technical documentation ($N = 5$). Participants also had a background in engineering ($N = 4$), installation ($N = 3$), or IT ($N = 3$). One participant was an elevator maintenance student.

Test Methods We used a combination of conventional qualitative and participatory methods: thinking aloud, participant observation, and questionnaires. These methods are known to reveal the behavior and perception of users [43]. Due to the small size of our test group, the statistical results are only indicative. Instead of relying on statistical inference, the quantitative results of the questionnaire are interpreted through experts' remarks and verbal observations. This research is at the conceptual phase, and qualitative methods support this phase best.

First, we observed participants during the test. The participants were also instructed to think aloud while performing the task. They were instructed to explain what they were doing, what they liked and disliked, what they did not understand, and voice it when having a doubt. The sessions were also video recorded so that we could go back and check a detail if needed.

The participants were also asked to complete a questionnaire to find out attitudes, preferences, and opinions regarding each method, and to compare the methods. In addition to user background questions, the questionnaire contained 7-point Likert-scale questions and free-form questions. In addition, one part of the questionnaire included 7-point Likert-scale questions related to the use of smart glasses.

Fig. 8. Test setup: Elevator shaft and car roof simulation (left); lobby (right).

The collected data were recorded and stored in a company SharePoint database. Data storage and removal observed company guidelines and policies as described on the user consent forms. The authors grouped similar items together and analyzed the data.

Our study did not involve ethical issues that would require ethical review according to the local regulations. Therefore, the study was exempt from Institutional Review Board (IRB) review.

Test Setup Two test tasks were meant to be performed on an elevator car roof. Because the elevator car is located inside an elevator shaft, and only qualified personnel are allowed to enter the shaft, in our user study, the shaft and the elevator car roof were simulated with a pallet with a fixed plywood top in a meeting room. The test equipment was located on the pallet (see Fig. 8). This simulation allowed us to avoid any safety-related issues that might arise from working on actual equipment, yet the task mimicked performing the task in a real environment. The setup was explained to the test participants, and during the tests, when requested to enter the car roof, they stepped on the pallet. The third task was meant to be performed in a building lobby. We simulated the lobby in the same meeting room, using a monitor as the TV screen. Another pallet with plywood was used to simulate the fixing of the equipment behind the TV.

Two identical rooms with identical test setups were used in the user tests. A 2-hour timeslot was reserved for each testing session.

The comprehension of symbols was not within the scope of this study; therefore, a list of symbols used in the instructions was available for the participants during the test (see Fig. 9). Participants were encouraged to check the meaning of any symbols during the study and ask for help if they could not understand the meaning of symbols.

Each test participant conducted a series of three tasks. For each of the three tasks, we had three different versions of the instructions created according to the selected compression methods: conventional, minimalist, and visual manual. Therefore, a total of nine different instruction sets were used in the tests. We varied both the order of tasks and the order of compression methods, and different tasks were tested with different methods by different people. Each test participant used and evaluated all three of the methods during the user tests.

User Test Session Flow The user test sessions followed a predefined flow. First, as most of the participants came to the user tests directly from other work duties, they were offered some refreshments and asked to relax and orient to the user test. While each participant was having refreshments, we explained the scope and purpose of the user test, asked for research consent, and requested permission to record the session and take photographs.

Second, we introduced the test setups, devices, and tools used in the test (see Fig. 10). We explained how smart glasses function and fitted the display for the participant. In addition, we went through the symbols and their “cheat sheets” on the wall.

Third, we emphasized that we were testing technology and information compression methods, not the participant. We explained that it is

Fig. 9. Test participant examining the list of symbols used with the visual manual method.

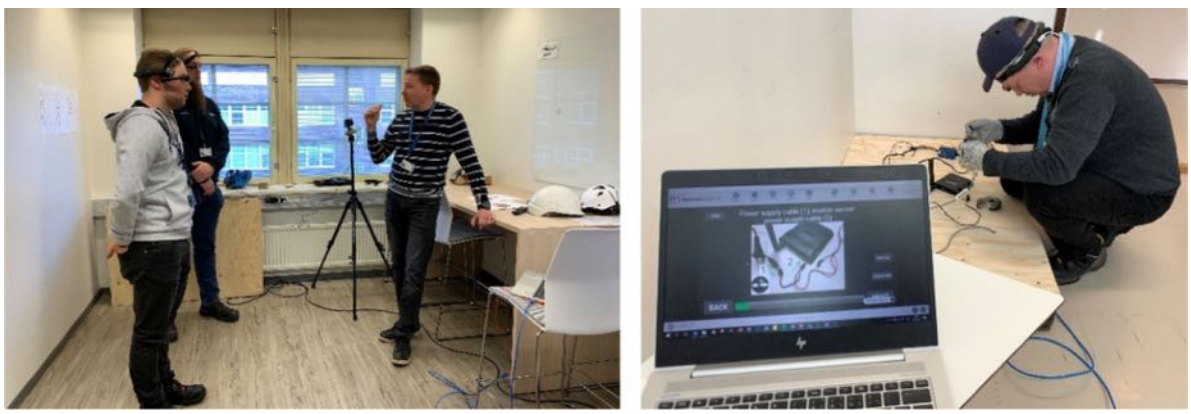


Fig. 10. Test setup and tasks being explained to participants (left). User test in progress (right). The smart glasses were connected to a PC so that we were able to follow what participants saw on the glasses.

important that the participants try not to please us, but openly express their thoughts and concerns. The participants were asked to think aloud, describe their actions, and express all the thoughts that they have regarding the tasks and the setup. They were also asked to think aloud when interpreting the instructions and experiencing any uncertainty.

When the actual user test began, participants performed the first task with the first method. After finishing the task, the participants answered questions regarding the first method. Two other tasks were completed, and questions were answered in the same manner.

After completing all three tasks, the participants had a short break, and were offered refreshments. The purpose of the break was to allow the participants to refresh their minds and get some distance from the previous task and method performed before answering the final part of the questionnaire. Finally, the participants answered questions comparing the methods as well as questions regarding smart glasses.

Results

First, all participants were very engaged with the user test. We did not measure user engagement per se, but we observed that all participants were focused, readily provided feedback, and showed intrinsic motivation during the tests [44]. Participants commented on other things than those primarily studied, and proposed improvements and modifications to components and products used in

TABLE III
PREFERENCE AND COMPREHENSION OF METHODS

Question	*N*	Result
Which version did you like the most?	20	Minimalist, *N* = 9 Visual manual, *N* = 7 Conventional, *N* = 4
Which version was the easiest to comprehend?	21	Minimalist, *N* = 10 Conventional, *N* = 6 Visual manual, *N* = 5

the test setup. They also gave feedback for the maintenance method used for the task and proposed alternative ways of performing the task. In addition, they discussed future maintenance guidance and proposed other areas where a similar system could bring benefits. All this confirmed high engagement and concentration from participants.

The preferred information design method among our test group was minimalist, followed by visual manual. The least liked method was conventional. Minimalist was rated the easiest to comprehend, conventional second, and visual manual last. The questions and results are presented in Table III.

As described earlier, the order of the tasks and methods was varied to avoid a bias that could be caused by participants favoring or disfavoring the method related to the task they performed first or last due to unfamiliarity with smart glasses or fatigue. This variability also addressed the bias caused by participants favoring or disfavoring a certain task.

There was a negative correlation between the method performed with the first task and the favorite method. Only two participants selected the method that they tried first as their favorite. Eight participants selected the last method as their favorite, and nine participants selected the middle one. One person omitted this question. There was also a negative correlation between the favorite method and the *Install motion sensor cable* task as only three participants preferred the one that they tested with the *Install motion sensor cable* task. Seven participants preferred the one they tested with the *Power-up Connection* task, and nine participants preferred the method they tested with the *Replace MediaPlayer* task.

Participants were requested to answer 12 questions for each method regarding the readability and understandability of the instructions on the screen, and the amount of text and number of graphics in the instructions. The results are presented in Figs. 11 and 12. In addition, participants were able to write free-form comments after each task regarding the task and the method.

Over 50% of participants agreed or totally agreed that there should be at least one graphic in each step, approximately 30% agreed or totally agreed that there should always be some text in each step, and approximately 30% agreed or totally agreed that there should always be both text and graphics in each step. Many participants commented that a combination of text and graphics complement each other especially with complex tasks: “There should always be an illustration for each step” (P1). However, with simple tasks it was not deemed necessary: “It is good that the very simple steps are not illustrated” (P2).

Participants were satisfied with the checklist type of reduced information; 48% disagreed or strongly disagreed with the question “I would like to have more explanatory text.” One third of participants were neutral, and only 9.5% of the participants agreed or strongly agreed. Participants also commented that the instructions described what was needed for task completion and that they did not need to assume or guess anything. Some participants named a specific task or detail for which they would have liked to have more explanation: “Where could I get more information about removing this part of the device?” (P3). These findings support our hypothesis of delivering reduced information and providing extra information only when the user needs and requests it.

With the visual manual instructions, people generally liked the concept but had problems recognizing when something needed to be done. For example, in deciding whether a graphic was a reference type of graphic only or whether they

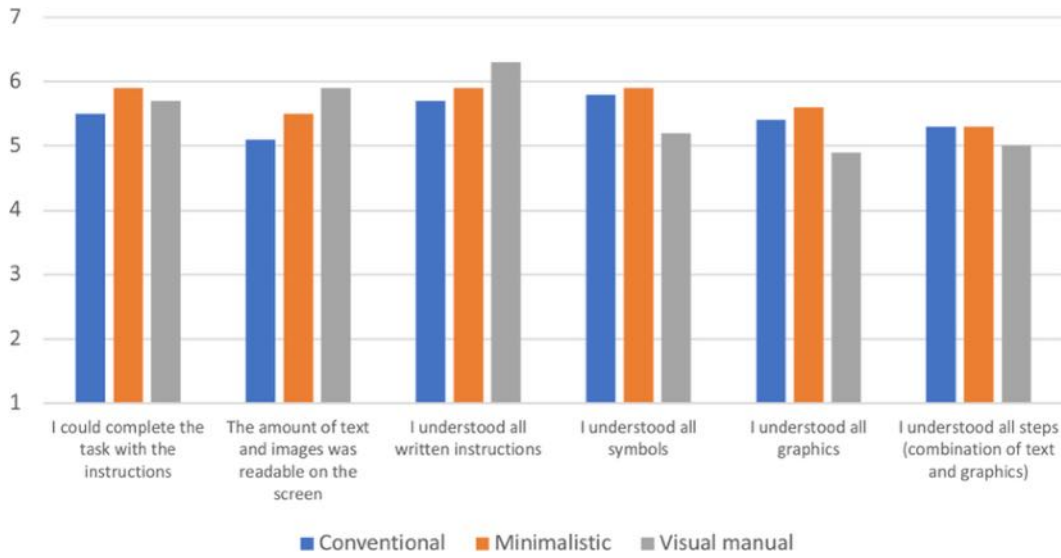


Fig. 11. Comments on methods, 1/2 (1 = totally disagree, 7 = totally agree), ($N = 21$). With these questions, the larger the value, the more positive the answer.

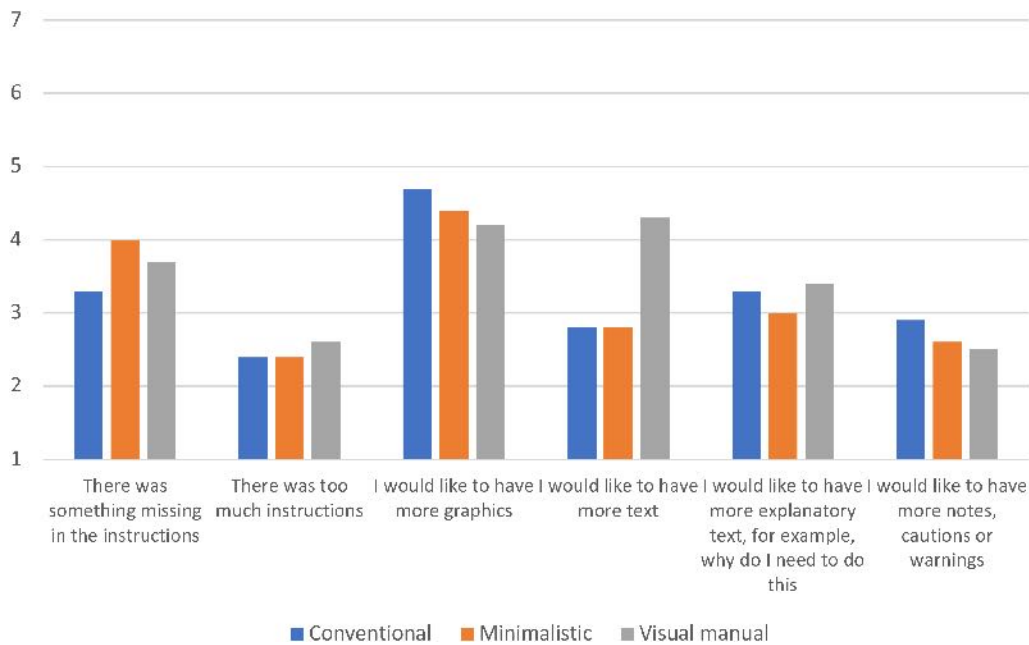


Fig. 12. Comments on methods, 2/2 (1 = totally disagree, 7 = totally agree), ($N = 21$). With these questions, the smaller the value, the more positive the answer.

needed to perform actions based on it, they might think aloud, "Do I need to do something now?" (P4) or "How do I know when I need to do something?" (P5). For one visual manual instruction (see Fig. 13), one screen consisted of a graphic that had several task symbols in it. The participants disliked the idea of several steps in one graphic and were also unsure whether the steps needed to be completed in a certain order: "There is so much information that I do not know where to start" (P1). This example revealed the need to clearly show one step at a time—that is, separated steps for *Drill hole* and *Route cable*.

Participants agreed that the amount of information was good for all the methods—that is, they were able to perform the tasks with the instructions provided. They felt that following the instructions one step at a time was more effective than scrolling up and down in a long list of steps that would not fit on the screen and would have required scrolling. Therefore, it can be noted that having the

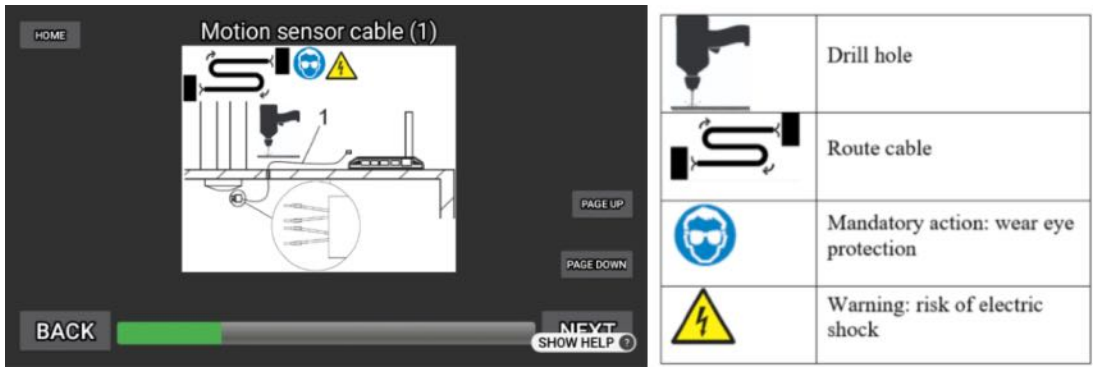


Fig. 13. Single graphic with several task symbols and legend from the visual manual (left). Symbols used in this instruction step (right).

TABLE IV
QUESTIONS REGARDING SMART CLASSES, ON A SCALE OF 1 = TOTALLY DISAGREE, 7 = TOTALLY AGREE (N = 21)

Question	Score	Rating Scale
The smart glass concept is something I could use in my daily work.	4.2	Neutral (3.4–4.5)
The use of smart glasses felt natural, did not hinder work.	4.2	Neutral
The use of smart glasses did not limit my awareness of the surroundings.	4.3	Neutral
The voice control operated well.	4.3	Neutral
The display was not in the way of my view.	4.8	Somewhat positive (4.6–5.7)
I was able to navigate and progress through the steps well.	5.3	Somewhat positive
The application UI was intuitive and easy to use.	5.6	Somewhat positive
The application worked well and ran smoothly.	4.7	Somewhat positive
I was able to read the text and graphics.	5.3	Somewhat positive
Instructions on smart glasses helps performing the task.	5.8	Very positive (5.8–7)

instructions proceed step-by-step on the screen was a well-received concept. However, the progress bar by itself was not enough to give a sense of location within the process for the participants. The relationship between steps and substeps was also not clear, and some participants even checked the next step to see what was coming before performing the step at hand: “I have a feeling that I need to check the next step here before doing anything” (P6).

Even though the list of the symbols used in the visual manual version was introduced to the participants before starting the task, and the participants were encouraged to check the meaning of symbols while doing the task, the visual manual approach scored the lowest for symbol understanding. Minimalism scored the highest in this category, followed by conventional. The conventional and minimalist versions used only symbols frequently included in maintenance instructions. Understanding all of the steps (a combination of text and graphics) followed this same logic: minimalist and conventional scored the highest, and the visual manual scored the lowest.

The visual manual scored high for the readability of text and graphics on the screen. The conventional manual scored the lowest, probably affected by the scrolling needed for some of the steps. Even though the conventional version had the most details and explanation, it scored the lowest for understanding all written instructions. The visual manual scored the highest for this category but admittedly had only a small amount of text. The minimalist scored above the conventional.

Smart glass use was generally received in a neutral or somewhat positive way (see Table IV). People were generally enthusiastic about the use of smart glasses, but as they are not yet a mature technology, that fact also showed in the evaluation

of the concept. For example, some participants had problems with the voice commands and had to repeat the same command several times, causing frustration. The small screen of the HMT-1 was also difficult to use for some participants, and they had problems either seeing the instructions properly or focusing their eyes on the small screen. Therefore, further development is needed for devices and their user interfaces to apply smart glass use to industrial maintenance. However, participants strongly agreed that the instructions displayed on smart glasses helped in performing the tasks. Several participants commented that the step-by-step instructions were easy to follow and helped them perform the steps in the correct order: "With these instructions, anybody could complete the task" (P7) and "The smart glasses seem very useful; I like the concept very much" (P8).

Discussion

The aim of this research was to better understand how information content should be created and designed to support future use in addition to existing delivery channels. Our focus was to adapt checklist-type of technical information to small screens such as smart phones, smart glasses, smart watches, and other wearables. The user tests were conducted using smart glasses.

XR offers exciting possibilities for industrial maintenance. However, as the needs of the users should be the driving force when creating the content for these solutions, the content creation cannot be automated, and information design has a central role in the development of content for XR solutions [45]. Furthermore, users are not a homogeneous group. Because the background and expertise of maintenance technicians vary, their needs also vary. Xue et al. note that to improve the usability of AR-assisted maintenance systems, the information should be contextualized, for example, according to the user's level of expertise and skills [18].

In our case study, we investigated the possibilities of delivering expert-level information through smart glasses. With the semantic tagging of DITA XML and competent information design, we could rely on certain elements containing the expert-level information and other elements containing the information needed for novices. Therefore, a straightforward filtering process of including and excluding information was utilized to contextualize the information according to a user's level of expertise. This design also catered to the reuse of information because the same XML files could be used in other applications or outputs to resolve the full content, including the content excluded in our test application. As no conditions were required in the design of our materials, we could utilize the existing XML files without conditional processing.

Even though it has been argued that users of technical instructions do not have to memorize procedure steps, many company style guides set limits on step lengths or promote chunking into meaningful pieces to reduce cognitive load. Especially with hands-busy tasks, such as elevator maintenance, the delivery of maintenance instructions has been problematic. Users might need to memorize a sequence of steps if they are unable to utilize the mobile device to scroll down lengthy step lists or flip the pages of a paper manual while they are performing the task.

With XR solutions, however, the number of steps or chunks in a procedure becomes irrelevant as the users are able to follow the instructions simultaneously while performing the task. In our user tests, we utilized a design that proceeded step by step, one screen at a time. The participants were able to control the application with voice commands synchronously with the task at hand. The attitude of participants toward displaying information this way in smart glasses was very positive, and it was deemed to assist in task completion. However, the information must be carefully designed so that there is only one action on each screen; otherwise, users might skip an action or be unsure of the order of the steps.

We used different methods to compress the information. The conventional version based on the original instructions was used as the baseline version in this study. The preferred method among our test group was minimalist, followed by the visual manual approach. The baseline, conventional, was the least-liked method in the study. Even though conventional had the most description and details, it was rated lower than the other two for task completion. Therefore, it can be concluded that more details do not necessarily aid understanding or task completion if those details are not seen as relevant or necessary by the user. In contrast, with minimalism, the information is designed with a user-centered approach, and the focus is on providing the correct amount of detail that users need to perform the task.

Researchers have suggested that leaving out text altogether might be an option for XR solutions [34].

The visual manual approach is an enticing theory, especially in a global setting because of resulting low translation costs. In our case study, the basic idea of the visual manual was evaluated positively, but participants felt that leaving out text altogether made understanding difficult, especially with more complex tasks. It should be noted that in industrial maintenance, working environments pose hazardous conditions, and the unambiguousness of instructions is essential. Previous research has indicated that text leaves less room for interpretation than graphics, and that language can be used as a disambiguation tool for graphics [46]. Furthermore, as the understanding of symbols was not within the scope of our study, more research is needed to evaluate the comprehension of action symbols, specifically in a global multicultural setting.

Participants were asked to evaluate whether each step should always have at least one graphic, some text, or both. There was a considerable difference among the answers to these questions. On one hand, participants preferred visual objects, and stated that they helped in understanding the location of connectors, for example. On the other hand, unnecessary details hindered task completion, so specific guidelines about the use of graphics and text should not be enforced. Again, the user-centered approach promoted by minimalism is the best guideline for the use of text and graphics. A fusion between minimalism and the visual manual method would be worth investigating in the future, and enriching the XML content with animations, videos, and augmented reality elements would also be an interesting area for further research.

Even though the testing was done with smart glasses, the results are also applicable to other devices with small screens, such as, for example, smart watches. Correctly tagged and compressed information enables the generation of relevant content to various outputs.

Furthermore, many test participants proposed that a similar system could be used, for example, in installation, document review in product development, or training. These suggestions further confirm that there are many uses for this type of information delivery.

Conclusion

In our study, we evaluated different information design and information compression methods and the usability of maintenance guidance on small screens such as smart glasses.

We wanted to study whether users of technical instructions are ready to accept smart glasses and XR as a delivery channel, and our research confirms that instructions displayed on smart glasses help in performing tasks in industrial maintenance, and attitudes toward using smart glasses are positive.

In this case study, we also explored whether the same information content can be utilized to deliver relevant content to the users based on their skill levels. Filtering information with different DITA XML elements proved to be an efficient way to scale information for different user needs. In this study, we used this method to provide checklist type of information for expert users and leave out novice-level information. In the future, we will continue this study by implementing the possibility to get further instructions when needed to cater more to novice-level users.

Our third area of interest in this study was the design of information content so that it supports use in smart glasses and other small-screen devices in addition to existing delivery channels. Because smart glasses as a delivery channel for technical instructions is a new concept, literature is lacking in guidelines for the design and authoring of technical instructions targeted for small screens. Information created using the minimalism principles was the preferred information design method within our test group. It was also considered the most understandable. Therefore, the minimalism heuristics could be implemented more thoroughly in the future. The visual manual concept was also received well and, in general, the participants of our study wanted to have more visual content. However, the cultural understanding of symbols needs to be further studied before implementation.

Research papers very often focus on augmented reality instructions for maintenance guidance—that is, systems where the actions to be performed are animated on top of a machine. However, for a large number of maintenance tasks, the necessary information can be presented with graphics or text, and using 3-D augmentations would be overengineering. Thus, it is important to study ways to utilize more traditional instructions, based on graphics and text, in hands-busy situations. On the other hand, in some use cases, animations and augmentations would be beneficial

and would bring extra value. Therefore, our proposal and area of further research is to embed augmented reality content in the XML files so that the viewer application can switch to the AR mode when the augmented content is available. In such a scenario, augmented content would be created only for complicated tasks, where the superimposed animations would help in task completion or accuracy, and the rest of the content would be derived from DITA XML files.

For industrial companies with tens of thousands of technical instruction sets, the capability to use the single source for multiple delivery formats without the need for constant tailoring would be a huge benefit. Our case study shows that well-designed technical information can be automatically used and delivered for several use cases and in several end devices, including displaying checklist-type maintenance information on smart glasses.

APPENDIX
QUESTIONNAIRE

1. Gender
   a. Male
   b. Female
   c. Other/I don't want to tell
2. Age
   a. < 20
   b. 20-29
   c. 30-39
   d. 40-49
   e. 50-59
   f. > 59
3. Your working experience related to maintenance field work
   a. No experience
   b. Very limited experience of field environment
   c. Somewhat familiar with field environment
   d. Familiar with field environment
   e. I have worked at field myself / very experienced
4. Field of expertise (select all that apply)
   a. Maintenance
   b. Technical Documentation
   c. Installation
   d. IT
   e. XR
   f. Engineering
   g. Student
   h. Other
5. Task/theory [participants chose the combination they had been assigned and rated statements/answered questions 6-9 separately for each combination]
   a. Install motion sensor cable
   b. Power up KONE Connection
   c. Replace KONE MediaPlayer

   —
   a. Conventional
   b. Visual manual
   c. Minimalist
6. Content [Likert scale for each question 1-8, 1 = strongly disagree, 7 = strongly agree, 8 = no opinion/cannot evaluate]
   a. I could complete the task with the instructions.
   b. There was something missing in the instructions.
   c. The amount of text and images was readable on the screen.

   d. There was too much instructions.
   e. I understood all written instructions.
   f. I understood all symbols.
   g. I understood all graphics.
   h. I understood all steps (combination of text and graphics).
   i. I would like to have more graphics.
   j. I would like to have more text.
   k. I would like to have more explanatory text, for example, why do I need to do this.
   l. I would like to have more notes, cautions or warnings.
   m. All steps should have at least some text.
7. Was there something that could be left out? [free-form field]
8. Was there something that was missing? [free-form field]
9. Do you have any other comments of feedback? [free-form field]
10. Which version did you like the most?
   a. Conventional
   b. Visual manual
   c. Minimalist
11. Why? [free-form field]
12. Which version was the easiest to comprehend?
   a. Conventional
   b. Visual manual
   c. Minimalist
13. Why? [free-form field]
14. Images and text: [Likert scale for each question 1-8, 1 = strongly disagree, 7 = strongly agree, 8 = no opinion/cannot evaluate]
   a. There should be at least one image at each step
   b. There should always be some text at each step
   c. There should always be both text and images at each step
15. Smart glasses: [Likert scale for each question 1-8, 1 = strongly disagree, 7 = strongly agree, 8 = no opinion/cannot evaluate]
   a. The smart glass concept is something I could use in my daily work.
   b. The use of smart glasses felt natural, did not hinder work.
   c. The use of smart glasses did not limit my awareness of the surroundings.
   d. The display was on the way of my view.
   e. The voice control operated well.
   f. I was able to navigate and progress through the steps well.
   g. The application UI was intuitive and easy to use.
   h. The application worked well and ran smoothly.
   i. I was able to read the text and images.
   j. Instructions on smart glasses helps performing the task.

References

[1] J. Tham, "Interactivity in an age of immersive media: Seven dimensions for wearable technology, internet of things, and technical communication," *Tech. Commun.*, vol. 65. no. 1, pp. 46–65, Feb. 2018. [Online]. Available: https://www.ingentaconnect.com/contentone/stc/tc/2018/00000065/00000001/art00006#:~:text=Seven%20dimensions%20of%20interactivity%20can,%2C%20and%207)%20sensory%20stimulation%2F

[2] J. Wolfartsberger, "Analyzing the potential of virtual reality for engineering design review," *Autom. Construct.*, vol. 104, pp. 27–37, Aug. 2019, doi: 10.1016/j.autcon.2019.03.018.

[3] T. Moser, M. Hohlagschwandtner, G. Kormann-Hainzl, S. Pölzlbauer, and J. Wolfartsberger, "Mixed reality applications in industry: Challenges and research areas," in *Proc. 11th Int. Conf.*, 2019, pp. 95–105, doi: 10.1007/978-3-030-05767-1_7.

[4] E. Kaasinen *et al.*, "Mobile service technician 4.0: Knowledge-sharing solutions for industrial field maintenance," *J. Interact. Des. Architecture*, no. 38, pp. 6–27, Autumn 2018. [Online]. Available: http://www.mifav.uniroma2.it/inevent/events/idea2010/doc/38_1.pdf

[5] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, "A systematic review of augmented reality applications in maintenance," *Robot. Cim-Int. Manuf.*, vol. 49, pp. 215–228, Feb. 2018, doi: 10.1016/j.rcim.2017.06.002.

[6] S. Siltanen and H. Heinonen, "Scalable and responsive information for industrial maintenance work—Developing XR support on smart glasses for maintenance technicians," in *Proc. 23rd Int. Conf. Academic Mindtrek*, 2020, pp. 100–109, doi: 10.1145/3377290.3377296.
[7] L. Bellamy, M. Carey, and J. Schlotfeldt, *DITA Best Practices: A Roadmap For Writing, Editing, and Architecting in DITA*. Upper Saddle River, NJ, USA: IBM Press, 2012.
[8] J. Tham, A. H. Duin, L. Gee, E. N., B. Abdelqader, and M. McGrath, "Understanding virtual reality: Presence, embodiment, and professional practice," *IEEE Trans. Profess. Commun.*, vol. 61, no. 2, pp. 178–195, Jun. 2018, doi: 10.1109/TPC.2018.2804238.
[9] M. Lorenz, S. Knopp, and P. Klimant, "Industrial augmented reality: Requirements for an augmented reality maintenance worker support system," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct*, Munich, Germany, 2018, pp. 151–153, doi: 10.1109/ISMAR-Adjunct.2018.00055.
[10] KONE Corporation - Improving the Flow of Urban Life, 2021. [Online]. Available: www.kone.com
[11] C. Niemöller, D. Metzger, L. Berkemeier, B. Zobel, and O. Thomas, "Mobile service support based on smart glasses," *J. Inf. Technol. Theory Appl.*, vol. 20, no. 1, pp. 77–108, Aug. 2019. [Online]. Available: https://aisel.aisnet.org/jitta/vol20/iss1/4
[12] A. Mewes, B. Hensen, F. Wacker, and C. Hansen, "Touchless interaction with software in interventional radiology and surgery: A systematic literature review," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 291–305, Feb. 2017, doi: 10.1007/s11548-016-1480-6.
[13] K. Klinker *et al.*, "Structure for innovations: A use case taxonomy for smart glasses in service processes," in *Proc. Multikonferenz Wirtschaftsinformatik*, 2018. [Online]. Available: https://www.researchgate.net/publication/321488897_Structure_for_innovations_A_use_case_taxonomy_for_smart_glasses_in_service_processes
[14] J. L. Aitken and H. A. Ross, "Mixed reality training method: Performance benefits for routine vehicle maintenance tasks," vol. 436 of DTA report, Defence Technology Agency, Feb. 2019. [Online]. Available: https://www.dta.mil.nz/assets/publication/0cc03bf296/Mixed-Reality-For-Vehicle-Maintenance-finaldoc.pdf
[15] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt, "A benchmark for interactive augmented reality instructions for assembly tasks," in *Proc. 14th Int. Conf. Mobile Ubiquitous Multimedia*, 2015, pp. 253–257, doi: 10.1145/2836041.2836067.
[16] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnulo, G. L. Cascella, and M. Fiorentino, "Evaluating the effectiveness of spatial augmented reality in smart manufacturing: A solution for manual working stations," *Int. J. Adv. Manuf. Technol.*, vol. 94, no. 1, pp. 509–521, Aug. 2018, doi: 10.1007/s00170-017-0846-4.
[17] S. Henderson and S. Feiner, "Exploring the benefits of augmented reality documentation for maintenance and repair," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 10, pp. 1355–1368, Oct. 2011, doi: 10.1109/TVCG.2010.245.
[18] H. Xue, P. Sharma, and F. Wild, "User satisfaction in augmented reality-based training using Microsoft hololens," *Computer*, vol. 8, no. 1, Jan. 2019, Art. no. 9, doi: 10.3390/computers8010009.
[19] SightCall—AR powered visual support, 2021. [Online]. Available: https://sightcall.com/
[20] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche, "An augmented reality training platform for assembly and maintenance skills," *Robot. Auton. Syst.*, vol. 61, no. 4, pp. 398–403, Apr. 2013, doi: 10.1016/j.robot.2012.09.013.
[21] S. Aromaa, I. Aaltonen, E. Kaasinen, J. Elo, and I. Parkkinen, "Use of wearable and augmented reality technologies in industrial maintenance work," in *Proc. 20th Int. Academic Mindtrek Conf.*, 2016, pp. 235–242, doi: 10.1145/2994310.2994321.
[22] S. Jayaprakash, *Technical Writing*. Mumbai, India: Himalaya, 2008.
[23] J. T. Hackos, *Content Management for Dynamic Web Delivery*. Hoboken, NJ, USA, USA: Wiley, 2002.
[24] M. Funk, T. Dingler, J. Cooper, and A. Schmid, "Stop helping me—I'm Bored! Why assembly assistance needs to be adaptive," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput./ACM Int. Symp. Wearable Comput.*, 2015, pp. 1269–1273, doi: 10.1145/2800835.2807942.
[25] K. V. Laan, *The Insider's Guide to Technical Writing*. Denver, CO, USA: XML Press, 2012.
[26] J. Redish, "Understanding readers," in *Techniques For Technical Communicators*, C. Barnum and S. Carliner, Eds. New York, NY, USA: Macmillan, 1993, pp. 14–41.
[27] K. Schriver, "What do technical communicators need to know about information design?," in *Solving Problems in Technical Communication*, J. Johnson-Eilola and S. A. Selber, Eds. Chicago, IL, USA: Univ. of Chicago Press, 2013, pp. 386–427.
[28] M. J. Kahn, K. C. Tan, and R. J. Beaton, "Reduction of cognitive workload through information chunking," in *Proc. Hum. Factors Soc. Annu. Meeting*, 1990, pp. 1509–1513, doi: 10.1177/154193129003401919.
[29] M. Coe, *Human Factors for Technical Communicators*. New York, USA: Wiley, 1996.
[30] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, no. 2, pp. 81–97, 1965, doi: 10.1037/h0043158.
[31] D. S. Sergio, *Tall Tales about the Mind and Brain: Separating Fact From Fiction*. New York, USA: Oxford Univ. Press, 2007.
[32] M. Gattullo, G. Scurati, M. Fiorentino, A. Uva, F. Ferrise, and M. Bordegoni, "Towards augmented reality manuals for industry 4.0: A methodology," *Robot. Comput.-Integr. Manuf.*, vol. 56, pp. 276–286, Apr. 2019, doi: 10.1016/j.rcim.2018.10.001.
[33] H. van der Meij and J. M. Carroll, "Principles and heuristics for designing minimalist instruction," *Tech. Commun.*, vol. 42, no. 2, pp. 243–261, May 1995. [Online]. Available: www.jstor.org/stable/43087895
[34] M. Gattullo, A. E. Uva, M. Fiorentino, G. W. Scurati, and F. Ferrise, "From paper manual to AR manual: Do we still need text?," *Proc. Manuf.*, vol. 11, pp. 1303–1310, Jun. 2017, doi: 10.1016/j.promfg.2017.07.258.

[35] G. W. Scurati, M. Gattullo, M. Fiorentino, F. Ferrise, M. Bordegoni, and A. E. Uva, "Converting maintenance actions into standard symbols for augmented reality applications in industry 4.0," *Comput. Ind.*, vol. 98, pp. 68–79, Jun. 2018, doi: 10.1016/j.compind.2018.02.001.
[36] C. Knopfle, J. Weidenhausen, L. Chauvigne, and I. Stock, "Template based authoring for AR based service scenarios," in *Proc. IEEE Virtual Reality*, Bonn, Germany, 2005, pp. 237–240, doi: 10.1109/VR.2005.1492779.
[37] I. Stock, M. Weber, and E. Steinmeier, "Metadata based authoring for technical documentation," in *Proc. 23rd Annu. Int. Conf. Des. Commun., Documenting Designing for Pervasive Inf.*, 2005, pp. 60–67, doi: 10.1145/1085313.1085330.
[38] T. Engelke, J. Keil, P. Rojtberg, F. Wientapper, S. Webel, and U. Bockholt, "Content first—A concept for industrial augmented reality maintenance applications using mobile devices," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Adelaide, SA, Australia, 2013, pp. 251–252, doi: 10.1109/ISMAR.2013.6671790.
[39] J. Gimeno, P. Morillo, J. M. Orduña, and M. Fernández, "A new AR authoring tool using depth maps for industrial procedures," *Comput. Ind.*, vol. 64, no. 9, pp. 1263–1271, Dec. 2013, doi: 10.1016/j.compind.2013.06.012.
[40] J. A. Erkoyuncu, I. F. del Amo, M. D. Mura, R. Roy, and G. Dini, "Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality," *CIRP Ann.*, vol. 66, no. 1, pp. 465–468, 2017, doi: 10.1016/j.cirp.2017.04.006.
[41] P. Ahola, "Instructions on small screens: Analysing the multimodality of technical communication through a design experiment," M.S. thesis, Dept. Modern Lang., Univ. of Helsinki, Helsinki, Finland, 2020. [Online]. Available: http://urn.fi/URN:NBN:fi:hulib-202009154094
[42] N. Harrison, "The Darwin information typing architecture (DITA): Applications for globalization," in *Proc. Int. Profess. Commun. Conf.*, 2005, pp. 115–121, doi: 10.1109/IPCC.2005.1494167.
[43] B. Kawulich, "Participant observation as a data collection method," *Forum: Qualitative Sozialforschung/Forum: Qualitative Social Res.*, vol. 6, no. 2, May 2005, doi: 10.17169/fqs-6.2.466.
[44] H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 938–955, Feb. 2008, doi: 10.1002/asi.20801.
[45] D. M. Armfield, A. H. Duin, and I. Pedersen, "Experiencing content: Heuristics for human-centered design for augmented reality," in *Proc. IEEE Int. Prof. Commun. Conf.*, Toronto, ON, Canada, 2018, pp. 241–247, doi: 10.1109/ProComm.2018.00057.
[46] M. Frixione and A. Lombardi, "Street signs and ikea instruction sheets: Pragmatics and pictorial communication," *Rev. Phil. Psych*, vol. 6, pp. 133–149, Oct. 2014, doi: 10.1007/s13164-014-0216-1.

Hanna Heinonen received the M.A. degree in English Translation and Interpretation from Tampere University, Tampere, Finland, in 2005, where she is currently working toward the Ph.D. degree in Interactive Technology. Since 2001, she has been working with technical communication, focusing on information architecture and design, user-centered, task-based documentation, and topic-based structured content creation. She is currently working as the Digital Content Lead at KONE Corporation. Her research interests include the role and potential of XR technologies in the field of technical communication, single-source publishing to different media and outputs, and user-centered documentation creation.

Sanni Siltanen received the M.Sc. degree in Applied Mathematics from Joensuu University, Joensuu, Finland, in 1999, the Lic.Sc. (Tech) degree in 2012 and the D.Sc. (Tech) degree in Information Sciences from Aalto University, Helsinki, Finland, in 2015. Her background is in XR research in the industrial context and she has been involved in numerous industry–academia jointly funded research projects. She is a Senior Expert in Research at KONE Corporation, Hyvinkää, Finland. Her current interests include artificial intelligence, mobile robotics, XR, sustainable urban flows, user interactions, human–computer interaction supporting industrial field work, and ways of enhancing industry-academia collaboration.

Petri Ahola received the B.A. and M.A. degrees in English Philology from the University of Helsinki, Helsinki, Finland, in 2018 and 2020, respectively. He is currently a Technical Communications Specialist with KONE Corporation, Hyvinkää, Finland. Since 2002, he has focused on user-centered technical content creation, while working with several major companies in Finland, including both IT and traditional industry. His current research interests include multimodality and single-source publishing to multiple channels, especially to small screens, AR and VR.

Research Article

Rediscovery of Developmental Research Articles in Electrical Engineering and Description of Their Macrostructure

—Gerald Rau

***Abstract*—Background:** *More than 30 years ago, Harmon distinguished developmental research articles (RAs), which propose a solution to a problem, from experimental RAs, but the developmental format has received little attention.* **Literature review:** *Genre analysis of RAs has been largely restricted to articles following the standard experimental/Introduction, Methods, Results, Discussion (IMRD) format, thereby excluding many developmental engineering articles. Recently, a textbook proposed Introduction, Process, Testing, Conclusion (IPTC) as a prototypical format for electrical engineering RAs, but this format has not yet been demonstrated from a corpus.* **Research questions:** *1. What is the macrostructure of electrical engineering RAs? 2. What are the characteristic features of each division of electrical engineering RAs?* **Methodology:** *Section headings, wordcount, and notable features were analyzed for 75 RAs from 15 electrical engineering journals and compared with both IPTC and Harmon's developmental structure.* **Results:** *Only one article, a case study, followed IMRD. Sixty-seven developmental RAs followed the IPTC format. These are distinguished by the second division (P), where the new solution is described, written in extended style, comprising several sections with headings specific to the research. A paragraph at the end of the Introduction describing the organization of the paper, the location of the theoretical framework and testing methods, and a ubiquitous Conclusion also differ from IMRD. Seven developmental RAs exhibited a hybrid format with the well-known IMRD section headings superimposed on an IPTC structure.* **Conclusions:** *Most electrical engineering articles are developmental and follow IPTC format. This can inform future genre analysis research and has pedagogical implications for teaching engineering writing.*

Index Terms*—Electrical engineering (EE), introduction, process, testing, conclusion (IPTC), genre analysis, macrostructure, research articles (RAs).*

Over 30 years ago, Harmon [1] described the "topical structure" of "experimental" papers in science. This structure, which he claimed had been codified in writing textbooks as early as 1927, was later called Introduction, Materials and Methods, Results, Discussion (IMRD) and remains the standard format for reporting experimental research. However, academic writing textbooks based on IMRD have proven inadequate for engineering students [2].

In the same article, Harmon described two other formats for engineering articles, "developmental" and "theoretical." Developmental articles appear to have received little attention, either in research or in writing textbooks, even though engineering articles typically report the development of a novel solution to a problem [3]. A new format—Introduction, Process, Testing, Conclusion (IPTC)—was recently proposed, with the claim that it is prototypical for electrical engineering (EE) research articles (RAs) [4]. This article will analyze 75 EE RAs to determine whether they follow IMRD or IPTC format, and whether they are experimental or developmental, and will describe some distinctive characteristics of the macrostructure of developmental articles. Following Rau [4], in this article, the term *section* is used in common sense for the named units designated by the author, and *division* for the conceptual units, either IMRD or IPTC.

The organization of the article is as follows. The Literature Review section shows that many RAs do not follow the IMRD format and, thus, have been excluded from previous genre analysis studies. The Methodology section discusses the difficulty of developing a representative corpus in EE, followed by the methods used in this study. The Results and Discussion section shows that all but one of the

Manuscript received February 12, 2021; revised July 21, 2021; accepted August 7, 2021. Date of publication October 25, 2021; date of current version November 24, 2021. *(Corresponding author: Gerald Rau.)*
The author is with the National Chung Cheng University, Chiayi 621301, Taiwan (email: gerryrau@ccu.edu.tw).
This article has supplementary downloadable material available at https://doi.org/10.1109/TPC.2021.3110618 provided by the author. The file consists of two appendices (164 kB in size).

IEEE 10.1109/TPC.2021.3110618

Practitioner Takeaway

- Almost all research articles in electrical engineering report the development of a new solution to a problem, following IPTC format rather than IMRD.
- IPTC format is characterized by a lengthy Process division describing the new solution, often comprising multiple sections with headings specific to the research, and written in extended style.
- Authors occasionally seem to impose the well-known IMRD section headings on work that otherwise follows the IPTC structure, a practice that may be confusing.

articles examined are developmental and follow the IPTC format. It also describes how this format differs from the well-known IMRD. Pedagogical implications are discussed, followed by a Conclusion.

Literature Review

Many genres are recognized within engineering and related fields [5], but this study focuses exclusively on the RA, a full-length primary research paper in a peer-reviewed journal. Much of the work on RA structure has been done within the genre analysis framework. Variants of genre analysis include New Rhetoric, Systemic Functional Linguistics, and English for Specific Purposes (ESP) [6]. In this section, I first review research in the ESP tradition that shows the variation in macrostructure of engineering RAs. Next, I demonstrate that many previous studies have been limited to the IMRD format. Then, I examine how the assumption that engineering is an applied science may have prevented recognition of the developmental format. The section concludes with the research questions.

Variation in Engineering Research Article Macrostructure Harmon, after a brief historical introduction, devoted most of his paper to a description of "experimental" RAs following a "topical structure," including the "heading [title and byline], abstract, introduction, methods or experimental details, results, discussion, conclusion, and acknowledgements and references" [1, p. 132]. This is the standard format of RAs which has come to be known as IMRD or Introduction, Materials and Methods, Results and Discussion, Conclusion (IMRaDC).

Nevertheless, research shows that the canonical IMRD structure is not always followed. Nonstandard headings are common in secondary RAs in applied linguistics [7] and can distinguish research methodologies in economics [8]. In information systems, RAs based on the hypothetico/deductive method of science tend to follow IMRD, but those using the problem/solution method that characterizes engineering do not [9]. Moreover, Posteguillo found that in place of a Methods section, RAs in computer science contain one or more sections presenting an "explanation of an algorithm or the process of implementing a system, program, or application" [10, p. 153].

Lin and Evans [11] examined the section headings of RAs in 39 disciplines and found an average of 7 variants of the IMRD structure for each of the 7 engineering and computing disciplines in the study. These included the presence of a Literature Review, use of a combined Results and Discussion, and addition of a Conclusion, variants that previous authors had often subsumed under the four common section headings. They noted that

> One consequence of ESP scholars' (conscious or subconscious) adoption of the "canonical" IMRD model is that it has limited the scope of some of their analyses. [11, p. 151]

Even so, from an initial sample of 20 RAs per discipline, they analyzed only "experimental" papers, defined as

> RAs containing the heading Method (or variations on this theme, such as "Experimental," "Empirical design," "The study," and "Data and research design"). (p. 152)

Only 12 of 20 papers in electronic and information engineering and 11 of 20 papers in EE fit this description (p. 159), meaning that the actual variability was far higher than their data show. They justify this fact based on distinguishing experimental from theoretical and review articles, but their designation of papers as experimental was based on section headings rather than

content, which I will later argue is an incorrect method of distinguishing them.

A related issue is that researchers often speak of engineering as a single discipline, although they divide science into multiple disciplines. For example, the British Academic Written English corpus does not distinguish between types of engineering at the university level, although it does distinguish between branches of science [12]. Similarly, Kanoksilapatham [13] called engineering a single discipline with biomedical, civil, and software as "subdisciplines," yet distinguished biochemistry and microbiology as separate disciplines. I will treat specialties commonly found in one academic department as subdisciplines within a discipline.

This assumption that engineering is a monolithic field has led researchers to expect engineering RAs to be more similar than they are. In concluding their study on five engineering fields, Maswana, Kanamaru, and Tajino stated,

> On its surface, engineering seems to be a unified discipline relating to the application of scientific principles and the production of useful things. However, it includes diverse subdisciplines ranging from observational experimentation to mathematical simulation, which was reflected in the diversity of the results. [14, p. 9]

Restriction of Studies to Experimental RAs

Many genre analysis studies have been restricted to articles where the IMRD format could be identified and to sections bearing those four names or closely related terms [9]. To date, only three studies appear to have examined complete RAs in EE, with only one considering non-IMRD articles.

Maswana et al. [14] compared articles from electrical, chemical, environmental, and structural engineering and from computer science. They found that RAs in these disciplines almost always included an Introduction and Conclusion. Everything between these they called the body (see [15]) since the other common IMRD section names were absent from many articles. Although their study contained many useful observations about the degree of variation, methodological issues limit its generalizability, particularly the decision to have one specialist in each subject area choose articles related to his or her own research, as will be discussed in the section on journal selection.

TABLE I
NUMBER AND PERCENT OF ARTICLES CONTAINING CLEARLY IDENTIFIED IMRD SECTIONS, OUT OF 60 RAS PER DISCIPLINE ANALYZED [DATA FROM 13]

	I		M		R		D	
	n	%	n	%	n	%	n	%
Biomedical	59	98	53	88	60	100	37	62
Civil	60	100	32	53	18	30	16	27
Software	60	100	16	27	27	45	39	65

Cotos, Huffman, and Link [16] conducted a massive study of 900 RAs from 30 research fields including EE, in an attempt to develop an interdisciplinary model of the move/step structure of each IMRD section. Thus, their first selection criterion was that each RA in the corpus had to follow the IMRD structure.

Kanoksilapatham, using the same corpus as her earlier studies on the Introduction [17] and Discussion [18], examined "the four macro sections of RAs" [13, p. 77] in biomedical, civil, and software engineering. For each field, 60 articles were selected, but moves were examined only in sections that were "independent and clearly marked" (p. 79) representatives of the four canonical IMRD sections. The number of articles fitting this pattern for each discipline are shown in Table I. Biomedical engineering (BME) was the closest to the expected IMRD format, with most RAs having the first three sections, but only 62% with a separate Discussion. For civil and software engineering, all articles had a clear Introduction but, on average, each of the other three sections occurred in fewer than half the articles. No explanation was provided for why they might be missing or what might be included in other sections that did not match the expected IMRD designations.

Studies of single divisions have also been limited to IMRD. The Introduction, the most studied of the four divisions, appears to be nearly ubiquitous (although occasionally with a different name or with no title) and has the most similar structure across fields. For many engineering fields, including civil, software, and BME [17], [19] and computer science [20], Swales' Creating a Research Space model [21, pp. 140–141] adequately describes the expected structure of the Introduction.

The Materials and Methods division has received the least attention. Cotos, Huffman, and Link [22], who provide the only description of a single division to include EE, used the same corpus as their

previous study [16], restricted to IMRD. Similarly, a study of Methods in eight science and social science disciplines looked only at "empirical data-driven RAs with an explicit Introduction-Method-Results-Discussion format" [23, p. 104].

Studies of the later divisions have been plagued by difficulties due to the different terms used, with Results, Discussion, Results and Discussion, Conclusion, and other section names such as Pedagogical Implications showing overlapping moves and steps (e.g., [16], [24]). This problem will be discussed further in the section on content and features of each division.

Developmental RA Macrostructure In addition to "experimental" papers in engineering, Harmon [1] also proposed "developmental" and "theoretical" paper formats. However, he devoted only one paragraph to the description of "developmental" papers, which report "development of a device, material, system, process, or method" (p. 136), although this description represents most papers in engineering.

Recognition of developmental papers as a separate format may have been hindered by the common view of engineering as applied science. For example, a genre analysis of university lab reports calls engineering "applied science" [12, p. 3]. This approach leads to the faulty assumption that reporting in science and engineering should be identical. Engineering has recently been incorporated into the next generation science standards (NGSS) in the US, creating a need to educate science teachers about differences between the nature of science and the nature of engineering. This approach led to the following definitions of science and engineering, based on the NGSS.

> **Science** involves asking questions and constructing explanations, using a systematic approach to develop models, carry out investigations, analyze and interpret data, and argue from evidence to understand the natural world.
>
> **Engineering** involves defining problems and finding solutions, using a systematic and often iterative approach to design products, processes, and systems to meet human needs and wants [25, p. 162].

This definition of engineering corresponds closely with Harmon's developmental paper which "reports on the development of a device, material, system, process, or method" in which "the introduction typically defines the problem to be solved" [1, p. 136]. Thus, although engineering is based on scientific principles, its goals and methods are totally different, so it should be expected that the format for writing up the research results will likewise be different. Moreover, since the new design is the goal and major contribution of this article, one would expect the portion of the RA reporting this development to be expanded relative to the data and interpretations (Results and Discussion), which are the emphasis of science.

An academic writing textbook recently claimed that RAs reporting development of a novel solution to a problem, including most EE articles, typically follow IPTC, a format analogous to IMRD but with distinctive characteristics [4]. However, to date, the predominance of this format has not been demonstrated from a corpus, nor has the similarity between Harmon's developmental structure and IPTC been evaluated.

Research Questions

Accordingly, this study attempts to address the following research questions:

> **RQ1.** What is the macrostructure of EE RAs?
>
> **RQ2.** What features are prominent in each division of EE RAs?

The results will be compared with the previous descriptions of IPTC and developmental formats. This approach will allow us to generalize about the prototypical format for RAs in EE, which has important pedagogical implications for those teaching engineering writing.

Methodology

In the first phase of this study, a small corpus (25 articles from five journals) was analyzed to determine the typical structure of RAs in EE. To verify the results, a larger and broader corpus (50 articles from 10 journals) was analyzed in the same way.

Corpus

Corpus Purpose and Type: The type of corpus desired depends on the purpose of the research. A corpus can be broad, representing an entire field or even many fields, or very specialized, representing a single subfield. It can include only the top journals in a field or a range of journals. This research was carried out primarily for

pedagogical purposes rather than basic research, so the purpose was to develop a corpus that would be representative of a broad range of research in different subfields of EE.

However, selecting representative journals is difficult in EE, as most journals are highly specialized, making it impossible to get agreement among subject area specialists on which journals are the most important. IEEE ACCESS, a multidisciplinary open-access journal, is the only high-impact nonspecialized journal in the field. Crossover of EE with other fields makes the problem even more intractable. For example, a journal in robotics may include articles related to EE, mechanical engineering, computer science, and even neurocognition.

A striking example of the difficulty of choosing a representative corpus is found in Maswana et al. [14], who asked one subject area specialist in each field to choose articles related to their research. For EE, the journals comprised the two premier science journals in the world (*Nature, Science*), three journals more closely related to physics than EE (*Advanced Materials, Applied Physics A, Journal of Applied Physics*), and two journals of letters (IEEE ELECTRON DEVICE LETTERS, *Physical Review Letters*), a different genre with a shorter and simpler structure than full RAs.

The Institute of Electrical and Electronics Engineers (IEEE) is by far the largest and most influential publisher of EE research journals and should constitute the bulk of any EE corpus. Among the 175 journals published by IEEE, 91 entitled Transactions and 25 entitled Journal report original research to an academic or research audience. Many of these are among the highest ranked journals in the field [26]. Magazines (30 titles) and those with topical names like *Computer* are trade publications, discussing applications of current research to industry, and have a different purpose, audience, and format. Letters contain brief reports of rapidly changing fields.

Journal and RA Selection: The selection of the corpus for this research was based specifically on the needs of the department in which the author teaches, the EE department of a research university. Professors in six of the seven lab groups frequently publish in IEEE journals, while the computer engineering group publishes primarily in computer science journals. Conversely, many professors from computer science and information engineering publish regularly in IEEE journals. Drawing a clear line between EE and CS is difficult, with the overlapping computer engineering specialty sometimes housed in either department. (For the distinction between computer science and computer engineering, see https://www.mastersportal.com/articles/2797/computer-science-vs-computer-engineering-which-degree-to-study-in-2021.html.). Currently, almost all work in EE involves computer modeling, design, algorithms, or simulations.

Of 36 full-time professors in the EE department, 34 have published in one to nine IEEE TRANSACTIONS. The other two, both from computer engineering, publish extensively in cross-disciplinary, topic-specific journals in specialty fields such as robotics or optics. At least 10 professors publish primarily in IEEE TRANSACTIONS. In total, the department has published in at least 59 IEEE research journals, including 41 Transactions and 14 Journals. Although two members of the department work in BME, this crossover field was excluded from this article as it is often considered a separate discipline.

Thus, in the first phase of this study, five IEEE TRANSACTIONS, representing five EE subfields, were chosen for analysis. All had been chosen as target journals by students in the author's English Technical Writing course. For each journal selected, the five most recent RAs published at the time (July 2016) were downloaded and analyzed, for a total of 25 RAs. Other genres including Survey, Review, Overview of a Standard, Brief Report, or Letter were excluded, as these have a different purpose and structure than primary research RAs.

In the second phase, journal selection was again based on the needs of the author's department. Thus, 10 additional journals were selected from a list of journals in which department members have published, one or two from each of six EE subfields. These represent a wide range of EE topics from large-scale systems (power grid), through materials and devices (electronic devices, microwave antennas), to the micro-level (integrated circuits), as well as processes (communication systems) and artificial intelligence (computer vision).

Five IEEE journals were selected in this phase, along with five from other publishers. Any journal with an impact factor less than 1 was excluded, as were journals where more than half the articles were from fields other than EE. This decision excluded several publishers, including Springer and Wiley, which mainly publish specialized

TABLE II
JOURNALS AND ABBREVIATIONS USED IN THIS PAPER

Corpus	Abbrev.	Journal	Publisher
1	*CAD*	**Computer-Aided Design of Integrated Circuits and Systems*	IEEE
	EC	**Energy Conversion*	IEEE
	MTT	**Microwave Theory and Techniques*	IEEE
	PAMI	**Pattern Analysis and Machine Intelligence*	IEEE
	SG	**Smart Grid*	IEEE
2	*C*	**Communications*	IEEE
	CC	*Computer Communications*	Elsevier
	EPSR	*Electric Power Systems Research*	Elsevier
	ESA	*Expert Systems with Applications*	Elsevier
	IE	**Industrial Electronics*	IEEE
	MAP	*IET Microwaves, Antennas & Propagation*	IET
	PD	**Power Delivery*	IEEE
	PR	*Pattern Recognition*	Elsevier
	SSC	*Journal of Solid States Circuits*	IEEE
	VLSI	**Very Large Scale Integration Systems*	IEEE

Note: *IEEE Transactions on …

crossover journals. For each journal, five of the most recent EE RAs accepted for publication at the time that the second phase began (January 2021) were downloaded, for a total of 50 RAs in this phase. Other genres and articles representing other disciplines in crossover journals were excluded. A list of the journals and abbreviations is shown in Table II, and the scope, subfield, and impact factors are listed in Table III. A list of the articles can be found in Appendix A in the supplementary material. The two corpora together will be called the combined corpus.

Description of RA Macrostructure in EE This study follows the basic procedure and nomenclature of Rau [4], where section designates the numbered and named units within an article, and division the conceptual units comprising the paper, either IMRD or IPTC. A division may include one or more sections; for example, the Introduction division of this article would include the unnamed initial section and the Literature Review section. In Rau [4], the designation of IPTC as a separate format was justified in part based on differences from IMRD in the number of sections, section headings, and division length. The first research question seeks to verify the overall structure of EE RAs based on the combined corpus.

First, the number of sections was recorded for each RA. Then, each section heading was designated as either general or specific. Some general section headings such as Introduction and Conclusion[s] are relatively invariant and conventional, but others show variation within a journal and greater variation from field to field. For example, in some fields, Materials and Methods may be called "Experimental Methods" (whether the methods would properly be classified as experimental or not). Similarly, a combined Results and Discussion may be called "Experiment," "Simulation," or "Validation." Other general headings include "Related Work" and "Preliminaries," common in computer science and related fields. On the other hand, in engineering, it is common in the second and sometimes the third division to select section headings specific to the content of the research being conducted, for example, "Techniques for Measurement of Magnetic Field Strength" or "Proposed Channel Feedback Algorithm." Titles that included both general and specific terms, such as "Results of 2-D FEM Optimization Procedure," were designated specific.

Next, each section of each article was assigned to a division, based on the section headings and general purpose of each section [4], as well as the position in the document, as it is important to consider the

TABLE III
SUBFIELDS AND IMPACT FACTORS OF THE JOURNALS IN THIS STUDY

Corpus	Abbrev.	Scope	Subfield*	Impact**
1	*CAD*	Computer-aided circuit design	IC	2.168
	EC	Energy conversion (motor, battery)	ED	4.501
	MTT	Microwave theory	MA	3.413
	PAMI	Computer image recognition	CV	17.861
	SG	Smart power grid	PG	8.267
2	*C*	Telecommunications	CS	5.646
	CC	Internet communications	CS	2.816
	EPSR	Power systems	PG	3.211
	ESA	Intelligent systems	CV	5.452
	IE	Electronics applications	ED	7.515
	MAP	Microwave antennas	MA	1.972
	PD	Power transmission	PG	3.681
	PR	Pattern recognition	CV	7.196
	SSC	Solid-state circuits	IC	4.929
	VLSI	Microelectronic systems	IC	2.037

Note: *CS = Communication systems, CV = Computer vision, ED = Electronic devices, IC = Integrated circuits, MA = Microwave antennas, PG = Power grid
Note: **Impact factor as listed on each journal homepage, January 2021

context when analyzing document structure (see Tseng [27] for the contextualization principle applied to assigning moves). Recall that engineering "involves defining problems and finding solutions" [25, p. 162], so the overall purpose or communicative function of each division in IPTC can be summarized as follows.

1. Introduction: Defining an important and unresolved problem
2. Process: Describing a solution and the reasons why it is workable
3. Testing: Showing that the solution is better than existing solutions
4. Conclusion: Summarizing the contribution of the current work

Thus, any section describing the reason for conducting the current research was assigned to the Introduction division. These could include initial sections without a heading (although no examples were seen in this corpus), as well as second sections that serve as a more extensive literature review, with the most common heading in engineering being "Related Work."

Sections describing the research being reported were assigned to the Process division, along with sections describing the theoretical basis necessary to understand it or the reasons why it is a reasonable and workable solution. The theoretical basis frequently precedes the description of the work and may have a heading like "Background," "Theoretical Foundation," "System Model," or variants. Sections with the heading "Problem Formulation" are ambiguous, as they may describe either a research gap (which would be part of the Introduction) or the framework that the current work is built on, so they require attention to the context and content of the section. Sections describing the current work frequently have specific headings related to the research goal, although they may also be more general but specific to the subfield, such as "Antenna Element."

Sections describing simulation or physical testing results were assigned to the Testing division, where the final design is evaluated and compared with other designs. Physical testing is often called "Experimental," although true experiments are rare in engineering. There is also some ambiguity with testing, as it is common in some subfields to test various portions of the design to establish optimum parameters prior to their compilation into the final design. In this work, such intermediate testing was considered part of the Process division, as it is

usually interleafed with the design description, and only the testing of the final design was considered to comprise the Testing division.

The Conclusion division is usually eponymously titled, sometimes with additional words. Occasionally, there may be a second section in this division specifying future work needed, although it is more common to include that information in a single Conclusion section.

After the sections were designated to divisions, the number of sections per division and the relative length of each division were compared. Wordcounts were obtained for each section, summed over a division, and standardized as a percent of the total article. Foxit Reader was used to obtain word counts directly from the PDF documents.

Mathematical expressions within a sentence were included in the wordcount, but numbered equations were excluded, as were tables, figures, lists of nomenclature, and algorithms. The extensive use of mathematical notations in many EE articles sometimes posed a problem in obtaining wordcounts, as different entry procedures will lead to different wordcounts for a single equation. For example, the equation $a^2 + b^2 = c^2$ might be considered five words based on spaces between terms or operators, a single word if entered from an equation editor, or no words if entered as a picture. Without extensive manual counts, this appears to be an intractable problem. For most articles, wordcount was based on spacing.

All coding was done by the author. Designation of titles as general or specific and assignment of sections to divisions are quite clear in more than 90% of the cases, so the accuracy of coding was not considered to be an important issue. All general section titles are shown in Appendix B in the supplementary material.

Description of Content and Features of Each Division of RAs in EE To answer the second research question, the text of each article was examined by the author to characterize the content and some of the key features of each division. Since there were no major differences between the two corpora in macrostructure analysis, they were considered together for this question.

Because this is the first study of this topic and it was conducted primarily to help engineering graduate students understand RA structure, emphasis was placed on identifying features that could be readily observed without the use of specialized tools such as concordance programs, multidimensional analysis (e.g., Jin [28], or detailed characterization of specific linguistic features such as speech acts (e.g., Sun, Wang, and Ren [29]) that have been used in more detailed research in other fields. Thus, this article reports general characteristics only, sufficient to show that IPTC and developmental represent the same format, with a more detailed study to follow.

First, the research goals were checked, to confirm whether each article represented a developmental RA [1]. Next, each division was examined for its general content and prominent features previously noted as characteristics of IPTC [4], including the use of bulleted lists to show the contributions of the article, an organizational paragraph at the end of the Introduction, summary paragraphs in the Process division that indicate the extended style, and recursion in the Testing division, with several components in each paragraph.

Results and Discussion

In the first subsection, I will answer the first research question, describing the macrostructure by determining the number of sections and the number of sections and words per division, and determining whether the section headings are general or specific to the research. In the second subsection, I will answer the second research question by characterizing the content of each division and determining whether the articles are developmental or experimental, based on their research goals. A discussion of pedagogical implications concludes the section.

Macrostructure of RAs in EE I will first present the number of sections and relative division length of the RAs, then the section names found in each division. Sixty-seven of the 75 articles match the expectations for IPTC. The final subsection describes the remaining eight articles in more detail.

Number of Sections and Division Length: Most RAs had five or six sections, but 30% had more and only two had four sections (see Table IV). It is interesting that most papers with more than six sections were in the first corpus. Articles in *PAMI* are on average the longest in the corpora, explaining the larger number of sections, but determining the reason for the greater number of sections in the others would require further study.

TABLE IV
TOTAL NUMBER OF SECTIONS IN EACH OF THE FIVE ARTICLES IN EACH JOURNAL

		Sections					
Corpus		**4**	**5**	**6**	**7**	**8**	**9**
1	*CAD*	-	-	2	2	1	-
	EC	-	-	3	-	2	-
	MTT	-	-	1	1	2	1
	PAMI	-	1	-	-	4	-
	SG	-	2	2	1	-	-
2	*C*	-	-	3	2	-	-
	CC	-	3	1	-	1	-
	EPSR	-	2	1	2	-	-
	ESA	-	3	1	1	-	-
	IE	-	3	2	-	-	-
	MAP	1	3	1	-	-	-
	PD	1	1	3	-	-	-
	PR	-	4	1	-	-	-
	SSC	-	2	2	-	1	-
	VLSI	-	2	1	1	1	-
Total		**2**	**26**	**24**	**10**	**12**	**1**
Mean				**6.1**			

The number of sections per division for each article was also determined, as shown in Table V. It is evident that most of the increase in the number of sections occurred in the second division.

The larger number of sections in the second division was reflected in the relative wordcount of the divisions. In IMRD format articles in the natural sciences, the Results and Discussion sections together typically represent more than half of the article [4]. However, in the journals in this study, the second division (Process) had the highest average wordcount in every journal except one, *PR*. On average, 48% of each article described the current design, a further indication of the importance of this division in developmental articles (see Table VI). The Testing division was longer than the Process in only 11 of the 75 RAs, including four in *EPSR*, three in *PR*, and only four in the other journals, so this difference may be related to the type of research reported.

Section Names: I next present a summary of the common section names in each division. A complete list of the general section names in the 75 articles is shown in Appendix B in the supplementary material.

Every article had an initial section called "Introduction." However, in 30% of the articles, a second section in the Introduction division provided more detail on past research. In many fields, this would be called "Literature Review," but in EE "Related Work(s)" is the most common heading. Almost all were found in journals in communication systems and computer vision, including four RAs in *CC* and all five in both *PAMI* and *PR*. These journals are closely related to computer science, where this section heading is nearly ubiquitous. "Related Work" also appeared as a subsection title within the Introduction section of two articles, both in journal *C*. This sort of overlap between section and subsection titles occurred in the other divisions as well. This work only considers the section headers, but it would be interesting to consider how many general headers are used in subsections as well.

Unlike the other three divisions, in the Process division, 75% of the section headings (137 of 182) were specific to the research being reported. General headings, where present, were almost always the first section in the Process division and were more common in certain journals, including *ESA* (8 of 9 sections in the division among the five RAs), *SG* (7 of 10), *MAP* (5 of 9), and *PR* (3 of 5). In two journals, every section heading in the division was specific (*EC*, *SSC*). Recurring general section headings in the Process division that presented the theoretical framework include "Background" (5, including 4 in *VLSI*), and various permutations of "Problem Formulation" (4, including 3 in *SG*). Headings of sections presenting an initial overview of the system included "System Model" (9). There were also sections that combined problem formulation and system description (3). Only eight articles out of 75 had a structure that might be considered IMRD, most with a section heading containing some variation of "Methods." These will be analyzed further in the next subsection.

Section headings in the Testing division were surprisingly diverse, although some journals appeared to have a preferred usage. In roughly two out of three articles, the headings were completely general, while others used a general term plus something specific to the research reported. Overall, the most common heading in this section was some variant on the word "Experiment," present in all five articles in both *PAMI* and *PR*, four articles in both *IE* and *MAP*, and three in *VLSI*. Computer testing was most reported in sections

TABLE V
NUMBER OF SECTIONS PER DIVISION IN EACH OF THE FIVE ARTICLES IN EACH JOURNAL

		I		P					T			C
Corpus		**1**	**2**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**1**
1	*CAD*	2	3	-	2	2	1	-	3	2	-	5
	EC	5	-	-	2	1	1	1	3	2	-	5
	MTT	5	-	-	1	2	2	-	1	3	1	5
	PAMI	-	5	1	-	-	4	-	5	-	-	5
	SG	3	2	-	5	-	-	-	4	1	-	5
2	*C*	4	1	-	1	2	2	-	5	-	-	5
	CC	2	3	2	2	1	-	-	4	1	-	5
	EPSR	5	-	-	3	2	-	-	3	1	1	5
	ESA	2	3	2	2	1	-	-	4	1	-	5
	IE	5	-	-	4	1	-	-	4	1	-	5
	MAP	5	-	2	2	1	-	-	4	1	-	5
	PD	5	-	2	1	2	-	-	5	-	-	5
	PR	-	5	5	-	-	-	-	4	1	-	5
	SSC	5	-	1	2	2	-	-	2	2	1	5
	VLSI	3	2	1	1	2	-	1	5	-	-	5
Total		**51**	**24**	**16**	**28**	**19**	**10**	**2**	**56**	**16**	**3**	**75**
Mean		**1.3**			**2.4**				**1.3**			**1**

including the word "Simulation," but there were many other alternatives (see Appendix B in the supplementary material).

Every article had a final section called "Conclusion," "Conclusions," "Conclusion and Future Work," or "Conclusion and Discussion," with one "Summery [sic] and Conclusion" and one "Final Remarks."

Articles With Intermediate Characteristics: Only eight articles might have been eligible for inclusion in previous genre analysis studies, as they have general section names corresponding to IMRD, especially Methods. The section headings and division length of these articles are shown in Table VII.

Of these eight papers, only one, *ESPR*1, was finally determined to be IMRD. Notably, this is the only article in this corpus which does not report the development of a new process or product. Rather, it is a case study that reports load data on the electric power system in Brazil and how it changed weekly during the COVID-19 pandemic. In the other seven articles, unlike experimental articles, Methodology sections described steps in the problem solution or design process, not a standard method followed by the researchers.

The paper *MAP*4 has only four sections, three of which could be variants of standard IMRD section headings. However, since the second division is both the longest and has a specific heading, it is better identified as IPTC. RAs in this journal, published by the Institution of Engineering and Technology (IET), were far shorter than those in any other journal, and this is the shortest in either corpus (1928 words), so the low number of sections number is undoubtedly related to that fact.

Apart from these, the other RAs in Table VII could be classified as hybrids, generally following the IPTC format, but using the common names of IMRD. The second article with four sections, *PD*2, is the only RA listed in this table published by IEEE and again is one of the shortest examined. The section headings are closely related to the standard headings of IMRD, but the second division is the longest, in line with the article's goal to "develop a modeling structure for ZnO surge arresters." Similarly, in the next three, *CC*2, *ESA*4, and *PR*2, the Process division is the longest division, as expected for IPTC. The authors of these articles seem to have been influenced by the prototypical IMRD pattern in their choice of section headings, although the content is more similar to IPTC.

Of the five articles from the journal *PR*, all except *PR*2 have a longer Testing division than Process

TABLE VI
AVERAGE LENGTH OF DIVISIONS IN EACH JOURNAL, AS A PERCENT OF A TOTAL

	CAD	*EC*	*MTT*	*PAMI*	*SG*
Average Words	8386	4882	8596	7041	6657
Division	%	%*	%	%*	%
I	19	13	15	19	18
P	51	56	56	44	56
T	28	25	23	34	23
C	2	5	6	2	3

	C	*CC*	*EPSR***	*ESA*	*IE*
Average Words	7592	7346	4805	9002	5420
Division	%*	%	%*	%*	%
I	24	28	20	25	17
P	52	43	38	41	52
T	21	26	37	29	27
C	4	3	4	4	4

	MAP	*PD*	*PR*	*SSC*	*VLSI*
Average Words	3007	5313	5771	5814	6611
Division	%	%	%*	%*	%*
I	21	16	30	9	18
P	49	45	30	56	53
T	24	33	35	33	26
C	6	6	4	3	2

Note: *Some columns do not sum to 100% due to rounding.
Note: **_EPSR_ excludes _EPSR1_ (IMRD format).

and also have a long Introduction, so the relative length of this division in *PR*1 seems to be a characteristic of this journal rather than related to the section headings. The section headings may again be a case of imposition of the typical IMRD format.

The final anomalous RA, *ESA*2 reports the combination of two existing designs into a single hybrid. This fact reduces the need for an extensive description of the design, leading to a shorter Process division. Otherwise, the article follows the IPTC format, although the section names are similar to IMRD.

Content and Key Features of Each Division of RAs in EE This section briefly describes some of the more obvious features of each division, with a goal of determining how they compare with the brief description of developmental articles by Harmon and the previous description of IPTC.

Introduction Division Content and Features: The Introduction division has a more similar structure across fields than any other division, moving from what is known from previous research to the research goal of the article. As with many RAs, the main goal of each article is usually found near the end of the Introduction [30]. Unlike IMRD, the goal is always presented as the aim, goal, or contribution of the paper, never as a research question. This fact reflects the difference between science, which "involves asking questions and constructing explanations," and engineering, which "involves defining problems and finding solutions" [25, p. 162], and is revealed in a problem-solution approach rather than moving from general to specific [31].

Although not selected for it, all but one article in the combined corpus met Harmon's criteria for "developmental" papers that report "the development of a device, material, system, process,

TABLE VII
EIGHT PAPERS WITH SECTION HEADINGS SIMILAR TO IMRD, WITH WORDCOUNT/DIVISION

	I	M/P	R&D/T	C
***EPSR*1**	1 Introduction	2 Materials and Methods 3 Statistical Analysis	4 Results and Discussions	5 Conclusions
	729	**646**	**2181**	**247**
***MAP*4**	1 Introduction	2 Antenna Element	3 Experimental Results	4 Conclusion
	592	**889**	**319**	**128**
***PD*2**	1 Introduction	2 Methodology	3 Model Usage, Data, and Results	4 Conclusions
	481	**2116**	**1736**	**379**
***CC*2**	1 Introduction 2 Related works	3 Methodology	4 Experiments	5 Conclusions
	3090	**2129**	**1302**	**219**
***ESA*4**	1 Introduction 2 Literature review	3 Principals [*sic*] and Theory 4 Methodology	5 Results and Analysis	6 Conclusion and Future Work
	2296	**2755**	**1704**	**458**
***PR*2**	1 Introduction 2 Related Work	3 Methodology	4 Results and Analysis	5 Conclusion
	2383	**2551**	**2110**	**380**
***PR*1**	1 Introduction 2 Related works	3 Methodology	4 Experiments	5 Conclusion
	1604	**1258**	**1429**	**157**
***ESA*2**	1 Introduction	2 Methodology Development	3 Case Studies 4 Results and Discussion	5 Conclusions
	1777	**1103**	**1963**	**211**

or method" [1, p. 136]. Although in my combined corpus, there were no reports of the development of new material, and it is not clear how Harmon intended to distinguish process from the method, 32 articles were classified as reporting a new process (approach, strategy, procedure, or algorithm), 20 reported a new device, and 7 reported a new system. A new model, a category perhaps representing either process or method in Harmon's list, was presented in 15 articles. Only the single case study was experimental, and none were purely theoretical.

Frequently, the main goal was marked with a phrase like "this paper" or "in this paper we" (42 occurrences), less commonly substituting "article" (11) or "work" (8). Altogether, this pattern was seen in 61 of 75 articles. The most common alternative (11) was the use of a statement of purpose followed by "we," as in "in order to … we" or "to overcome … we." The most common verbs were *propose* (32), *present* (15), or *develop* (6), with the verb in the passive voice in 22 articles. Almost half the papers (35 of 75) included a bulleted or numbered list of contributions near the end of the Introduction. Every RA in *C* and *CC*, both in the subfield communication systems, included such a list.

In 28 of 75 articles, multiple research goals were presented in successive paragraphs, alternating with the need for the work. An example is found in *MTT*3, where paragraph 3 says, "In order to reduce the complexity of a full size AAA architecture, we propose …." The next paragraph states, "The focus of this paper is to design …." Three paragraphs later, we find, "While doing so, we aim to answer some fundamental theoretical and practical questions such as the following: [3 bullets omitted]." The next subsection concludes, "The signal processing part of this paper aims to provide …." Finally, the last subsection lists "The specific contributions of the paper are as follows: [3 bullets omitted]." This structure, with multiple research goals split over several paragraphs, is not uncommon in EE, although this is an extreme example.

Similarly, in communication systems journals, it is common after reporting the limitations of previous research to summarize the new solution and how it

overcomes each limitation. Four of the five articles in *C* and two in *CC* employed this pattern, followed by a bulleted list of contributions, rather than a single research goal. With research on a system, it may be more difficult to specify a single goal, as all the parts must work together for the system to function. Overall, numbered lists of contributions were found in 11 articles, and bulleted lists in 17—together, more than one-third of the RAs. Nevertheless, although the research goal was often not clearly stated in a single sentence in the Introduction, in many of these articles, it was stated in a single sentence in the Abstract, in line with the idea that in engineering, the main contribution should be clear from the Title, Abstract, Introduction, and main text of the paper [3].

As mentioned in the previous section, in many articles, the Introduction division contained a second section, frequently called "Related Work(s)." In these articles, the research goal is often stated in the Introduction section and reiterated at the end of the second section. In one article (*CC*1), the Related Work section was moved to the end, just before the Conclusion. The reason for this location is not clear, but perhaps represents backgrounding, as when the Method section is moved to the end of some IMRD articles, since the material is familiar to experienced researchers, who will skip over it.

In 58 RAs (77%), the Introduction section concludes with a paragraph or part of a paragraph that briefly describes the organization of the work, summarizing the content of each section. Although common in linguistics, this feature is rare in science articles.

Process Division Content and Features: The Process division often begins with a section describing the theoretical foundation necessary to understand the present work. Although the theoretical foundation is often found in the Introduction division in IMRD format, in EE it is frequently found in the same section as the current design, often as the first subsection of the Process division. For example, in two articles where the second or third section is called Methodology, the first subsection is "Problem Formulation and Existing Models" (*PD*2) or "Problem Definition and Notations" (*PR*1). Sections called "Background" may contain a comparison of various models, leading to a justification for using the one chosen for the present study.

The largest part of the Process division, however, describes the new design of the product or process developed, with both section and subsection headings specific to the research. Furthermore, unlike IMRD, where the Methods were often written in condensed style [32], in IPTC the analogous Process division is written in extended style. This is evident from the extensive use of summary paragraphs at the beginning of a section before the subsections. Introductory summary paragraphs were found in 52 of 75 articles. The number and length of subsection headings differ by the journal. Although many journals have multiple subsections, with long descriptive headings, subsections are rare in *MAP*.

Testing Division Content and Features: Section headings in the testing division frequently include the word *experiment*. This may have misled some previous researchers to classify these articles as experimental when, in fact, they are developmental. In EE, the term experimental denotes physical testing as opposed to simulation testing. Rarely does the procedure meet the scientific definition of an experiment, which involves the manipulation of an experimental variable and measuring the effect on a response variable, usually with replication and control. The term is never used for computer testing, however. This fact may explain Kwan's observation that articles in information systems that follow the design systems research paradigm usually have sections entitled "Evaluation" rather than "Experiment" [9].

In IMRD, the testing methods, test results, and comparison with previous studies are usually found in three separate divisions (Methods, Results, and Discussion, respectively). In striking contrast, in IPTC, these are often found in a single paragraph. An example can be seen in the following paragraph from *MAP*3.

> 1. Finally, we measure the RCS performance of the fabricated prototype from 8.4 to 13.4 GHz. 2. The experimental results of monostatic RCS for normal incidence under different polarisations are plotted in Fig. 12f. 3. The black line indicates the RCS of a reference PEC plane with the same dimension as the proposed antenna, and the two dashed lines indicate the RCS of the proposed antenna for x- and y-polarisations. 4. The red and blue lines are the calculated monostatic RCS reductions for x- and y-polarisations. 5. The −7 dB bandwidth is from 9.4 to 13.7 GHz (37.23% relative bandwidth). 6. A significant deterioration occurs from 10.8 to 11.2 GHz, which is mainly caused by the large deviation between the actual phase difference and 180°

> phase difference. 7. The maximum RCS reduction reaches 21.05 dB at 12.1 GHz for x-polarisation and 21.62 dB at 12.2 GHz for y-polarisation. 8. It is obvious that the state of varactor diode has little effect on RCS reduction. [p. 1939, sentence numbers added]

In this paragraph, sentence 1 describes the basic test, sentence 2 points to the results in the figure, sentences 3-4 explain how to read the figure, and sentence 5 highlights the important information, followed by an explanation of the data in sentence 6. Sentence 7 then calls attention to another main point, with a comment on that point in sentence 8.

The same pattern (testing method, pointing to a graphic, highlighting data, comparison, or explanation) is often repeated recursively for several different tests. Paragraphs containing multiple components, leading to a recursive structure, were found in 70 of 75 of the RAs, showing that this is an expected feature in EE RAs.

Conclusion Division Content and Features: As noted above, every article in the corpus had a separate Conclusion. This may include a summary of the article (55 of 75 RAs), its contribution to the body of knowledge (47 of 75), and future work planned or in progress (26 of 75). In a few cases, the contribution appears as a numbered (3 of 75) or bulleted (2 of 75) list.

Comparison of IPTC With Harmon's "Developmental" Format: Based on the results presented above, the articles in my combined corpus selected from prominent EE journals do not fit the expectations of the standard IMRD format, but rather those of IPTC. The IPTC format was developed independently before seeing Harmon's work, and the description is far more detailed. Nevertheless, it matches very closely the format described by Harmon for developmental papers.

> Much of the above discussion about the structure of the experimental paper holds for the developmental paper. After the heading and abstract, the introduction typically *defines the problem to be solved* and the approach of the *author*, discusses *previous work* on the same problem or one close to it, and reviews any pertinent *theoretical principles* necessary for understanding how or why the new development works. The *main section lays out the essential features of this new development*, followed by a discussion of its *operating characteristics under various test conditions*, *advantages* over similar developments, and, if appropriate, costs. The concluding section *reiterates the novel features* of the development, reviews and interprets the test results, and points out *possible directions for future work* and applications. Acknowledgments and references are the same as for an experimental paper. [1, p. 136, emphasis added]

The RAs in this study incorporate all the elements highlighted, in the same order. The Introduction division defines the problem to be solved and discusses previous work. In the Process division, the theoretical principles underlying the work are presented, followed by the main section that lays out the essential features of this new development. The operating characteristics under various test conditions and advantages of the new design are shown in the Testing division. The Conclusion division summarizes the work, reiterating the contribution, and pointing to directions for future work.

The small number of articles selected from each journal made it impossible to determine whether deviations from the expected IPTC format, such as the articles where the Testing division was longer than the Process, were due to expectations of a journal, differences in the research, or the authors' writing style. Some articles appeared to be written in the order that the work was done, rather than being reorganized according to logical argument structure for ease of reading, a common problem among novice writers. There also appeared to be similarities in style among authors from the same country, a fact that may reflect the imposition of the writing standards of another language. Very nonstandard English was noted in three articles in *EC*, in line with previous observations that even the best papers in IEEE journals are often characterized by nonstandard grammar [33]. Similarly, graphics or references were out of order in at least eight of the RAs in the first corpus of 25 RAs, indicating that editing for such details may be less important in EE than in other disciplines.

Although none were identified in this study, engineering papers with a mathematical focus may have an Introduction, Proof, Conclusion structure. Since the contribution is logical proof, empirical testing is unnecessary [4]. This format seems to be equivalent to the "theoretical" papers proposed by Harmon [1], [34] and the format for pure mathematics described by Kuteeva and McGrath [35], but confirmation of that supposition awaits further study.

Pedagogical Implications

Graduate students learning to write research papers benefit from models derived from genre analysis but may have problems if the model does not adequately describe the structure of articles in their field [36]. Genre researchers are increasingly discovering that even within a field there may be more than one basic format, based on whether the work is primary or secondary research [7], and differences in research methodology (e.g., [8], [9]). I have also noted this fact in a preliminary study of civil engineering articles, where articles describing basic materials tended to follow IMRD, but those dealing with structures tended to follow IPTC.

Given the limited research on engineering RAs, it is not surprising that most textbooks on academic writing either do not discuss differences between science and engineering RAs or consider engineering as a minor variant on the IMRD model. For example, in the textbook *Writing for Science and Engineering* [37], the word "engineering" does not appear in the chapter on RAs. Textbooks on technical writing (e.g., [38], [39]) provide useful ideas about the writing process and characteristics of good technical writing in general, but contain only one chapter or even a single section on writing RAs.

One of the better textbooks on scientific writing, by Cargill and O'Connor [15], distinguishes an IBC (Introduction, Body, Conclusions) format common in "some kinds of engineering, computer science, remote sensing, and physics," in which the body sections "may contain various combinations of theory, methods, results, and discussion" (p. 15). Unfortunately, this is the only mention of IBC format or engineering in the textbook. To the best of my knowledge, the only textbook to clearly distinguish the differences between science and engineering RA structure is my own [4].

Based on the descriptions above, some main characteristics of IPTC (developmental RAs) that differ from IMRaDC (experimental RAs) are summarized in Table VIII. To allow easier comparison, divisions are designated by ordinal numbers as well as names, with R&D considered a single division. These features can be used by students for an initial determination of format, as this study showed that they are valid in approximately 90% of cases.

This study has obvious pedagogical implications for anyone teaching academic writing for engineering graduate students and even undergraduates. Explicit instruction in the macrostructure differences between IMRD and IPTC will help students understand the structure of articles in their field. Nevertheless, although most RAs in EE are developmental and follow IPTC format, there are major differences in expected article length and section headings from journal to journal. There are also journal-specific differences in details such as the formatting of section headings, tables, and figures. Thus, novice writers need to examine the structure of exemplars in their target journal rather than following a fixed pattern. Although a general pattern is helpful as a starting point, it will almost never be correct in all specifics.

TABLE VIII
Differences Between Experimental (IMR&DC) and Developmental (IPTC) Formats

	IMR&DC	IPTC
Number of sections	4-5	5-8
Names of sections	General	Specific for P
Longest division	3rd* (R&D)	2nd (P)
Theoretical framework location	1st (I)	2nd (P)
Testing methods location	2nd (M)	3rd (T)
2nd division style	Condensed	Extended

Note: Ordinal numbers indicate division.

Teachers should be aware that complications may occur in assigning sections to divisions. For example, in some EE articles, two distinct parts of a design are reported sequentially, each part in a section, with development and testing as subsections. In such cases, it may be necessary to assign subsections to divisions with a recurring P-T-P-T structure. As these atypical structures seem in my experience to be more common in lower quality journals, none were encountered in the present corpus, but they may be found in student exemplars.

Conclusion

Based on the research goals and characterization of each division, IPTC is not a totally new format, but a rediscovery of the developmental format proposed by Harmon [1]. In this article, the format was characterized in more detail, and examination of a corpus revealed that the vast majority of RAs in EE follow the IPTC format or a hybrid, the single exception being a case study that followed IMRD.

Some distinguishing characteristics of IPTC include a larger number of sections, with section headings

specific to the research particularly in the Process division, which is typically written in extended style rather than in the condensed style common in IMRD. This finding is logical, as the solution presented in that division is the main contribution of the article. The location of important components also differs from IMRD, with the theoretical framework often found in the beginning of the Process division and the testing methods usually found in the Testing division, in the same paragraph with the results and comments on the results.

Nevertheless, there is great variation among EE journals in terms of total article length, relative division length, name and number of sections, and organization. Since RAs do not follow a fixed pattern even within EE, writers need to analyze exemplar articles from their target journal to determine the expected format. This fact has pedagogical implications in teaching EE students how to write RAs.

This description of developmental/IPTC articles can inform future genre analysis of engineering RAs. Moreover, the designation of divisions avoids problems that previous authors have faced with varied section headings and may be useful in future work on IMRD RAs. This would also allow direct comparison of the IMRD and IPTC formats, which was not possible in this study since only one example of IMRD was found in the combined corpus examined.

Future work on the same combined corpus will use component analysis to analyze the argument structure of IPTC articles at a finer level. It will also show how different types of evidence are used in support of different component claims, and how the location of graphics and citations differs between IMRD and IPTC.

References

[1] J. E. Harmon, "The structure of scientific and engineering papers: A historical perspective," *IEEE Trans. Prof. Commun.*, vol. 32, no. 3, pp. 132–138, Sep. 1989.
[2] G. Rau, "Development of component analysis to support a research-based curriculum for writing engineering research articles," *English Specific Purpose*, vol. 62, pp. 46–57, 2021. [Online]. Available: https://doi.org 10.1016/j.esp.2020.12.001
[3] I. Stojmenović and V. Milutinović, "How to write research articles in computing and engineering disciplines," *Singidunum J. Appl. Sci.*, vol. 9, no. 1, pp. 42–50, 2012. [Online]. Available: https://doi.org/10.5937/sjas 1201042s
[4] G. Rau, *Writing for Engineering and Science Students: Staking Your Claim*. London, UK: Routledge, 2020.
[5] T. Orr, "Genre in the field of computer science and computer engineering," *IEEE Trans. Prof. Commun.*, vol. 42, no. 1, pp. 32–37, Mar. 1999.
[6] K. Hyland, "Genre and second language writing," in *The TESOL Encyclopedia of English Language Teaching*, J. I. Liontas Ed., 1st ed. Hoboken, NJ, USA: Wiley, 2018, pp. 2359–2364.
[7] R. Yang and D. Allison, "Research articles in applied linguistics: Structures from a functional perspective," *English Specific Purpose*, vol. 23, no. 3, pp. 264–279, 2004. [Online]. Available: https://doi.org/10.1016/s0889-4906(03)00005-x
[8] G. Jin, C. Li, and Y. Sun, "Exploring the macrostructure of research articles in economics," *IEEE Trans. Prof. Commun.*, vol. 63, no. 3, pp. 227–243, Sep. 2020.
[9] B. S. Kwan, "A cross-paradigm macro-structure analysis of research articles in information systems," *English Specific Purpose*, vol. 45, pp. 14–30, 2017. [Online]. Available: https://doi.org/10.1016/j.esp.2016.08.002
[10] S. Posteguillo, "The schematic structure of computer science research articles," *English Specific Purpose*, vol. 18, no. 2, pp. 139–160, 1999. [Online]. Available: https://doi.org/10.1016/s0889-4906(98)00001-5
[11] L. Lin and S. Evans, "Structural patterns in empirical research articles: A cross-disciplinary study," *English Specific Purpose*, vol. 31, no. 3, pp. 150–160, 2012. [Online]. Available: https://doi.org/10.1016/j.esp. 2011.10.002
[12] J. Parkinson, "The student laboratory report genre: A genre analysis," *English Specific Purpose*, vol. 45, pp. 1–13, 2017. [Online]. Available: https://doi.org/10.1016/j.esp.2016.08.001
[13] B. Kanoksilapatham, "Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora," *English Specific Purpose*, vol. 37, pp. 74–86, 2015. [Online]. Available: https://doi.org/10.1016/j.esp.2014.06.008
[14] S. Maswana, T. Kanamaru, and A. Tajino, "Move analysis of research articles across five engineering fields: What they share and what they do not," *Ampersand*, vol. 2, pp. 1–11, 2015. [Online]. Available: https://doi.org/10. 1016/j.amper.2014.12.002
[15] M. Cargill and P. O'Connor, *Writing Scientific Research Articles: Strategy and Steps*. West Sussex, UK: Wiley, 2013.
[16] E. Cotos, S. Huffman, and S. Link, "Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy," *J. English Acad. Purpose*, vol. 19, pp. 52–72, 2015. [Online]. Available: https://doi.org/10.1016/j.jeap.2015.05.004

[17] B. Kanoksilapatham, "Research article structure of research article introductions in three engineering subdisciplines," *IEEE Trans. Prof. Commun.*, vol. 55, no. 4, pp. 294–309, Dec. 2012.
[18] B. Kanoksilapatham, "In search of the generic identity of the discussion section: Three engineering sub-disciplines," *Taiwan Int. ESP J.*, vol. 4, no. 2, pp. 1–26, 2012.
[19] B. Kanoksilapatham, "Civil engineering research article Introductions: Textual structure and linguistic characterization," *Asian ESP J.*, vol. 7, no. 2, pp. 55–84, 2011.
[20] W. Shehzad, "Move two: Establishing a niche," *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos*, vol. 15, pp. 25–49, 2008.
[21] J. M. Swales, *Genre analysis: English in Academic and Research Settings*. Cambridge, UK: Cambridge University Press, 1990.
[22] E. Cotos, S. Huffman, and S. Link, "A move/step model for methods sections: Demonstrating rigour and credibility," *English Specific Purpose*, vol. 46, pp. 90–106, 2017. [Online]. Available: https://doi.org/10.1016/j.esp.2017.01.001
[23] M. Peacock, "The structure of the methods section in research articles across eight disciplines," *Asian ESP J.*, vol. 7, no. 2, pp. 99–124, 2011.
[24] R. Yang and D. Allison, "Research articles in applied linguistics: Moving from results to conclusions," *English Specific Purpose*, vol. 22, no. 4, pp. 365–385, 2003. [Online]. Available: https://doi.org/10.1016/s0889-4906(02)00026-1.
[25] G. Rau and A. Antink-Meyer, "Distinguishing between science, engineering and technology," in *Nature of Science in Science Instruction: Rationales and Strategies*, W. McComas Ed., (in Science: Philosophy, History and Education Ser.). Cham, Switzerland: Springer, 2020, ch. 8, pp. 159–176.
[26] IEEE. IEEE Journals Continue to Excel in Citation Rankings. [Online]. Available: https://innovate.ieee.org/ieee-journals-continue-to-excel-in-citation-rankings-2/.
[27] M.-Y. Tseng, "Creating a theoretical framework: On the move structure of theoretical framework sections in research articles related to language and linguistics," *J. English Acad. Purposes*, vol. 33, pp. 82–99, 2018. [Online]. Available: https://doi.org/10.1016/j.jeap.2018.01.002
[28] B. Jin, "A multidimensional analysis of research article discussion sections in the field of chemical engineering," *IEEE Trans. Prof. Commun.*, vol. 61, no. 3, pp. 242–256, Sep. 2018.
[29] Y. Sun, G. Wang, and H. Ren, "To entertain or to serve: Chinese and US Banks' online identity based on a genre analysis of social media," *IEEE Trans. Prof. Commun.*, vol. 64, no. 2, pp. 121–136, Jun. 2021.
[30] G. Myers, "In this paper we report …": Speech acts and scientific facts," *J. Pragmatics*, vol. 17, no. 4, pp. 295–313, 1992. [Online]. Available: https://doi.org/10.1016/0378-2166(92)90013-2
[31] L. L. Khaw and W. W. Tan, "Establishing a territory in the introductions of engineering research articles using a problem-solution patterns approach," *IEEE Trans. Prof. Commun.*, vol. 61, no. 2, pp. 133–150, Jun. 2018.
[32] J. M. Swales and C. B. Feak, *Academic Writing for Graduate Students: Essential Tasks and Skills*, vol. 1, Ann Arbor, MI, USA: Univ. Michigan Press, 2004.
[33] W. Rozycki and N. H. Johnson, "Non-canonical grammar in Best Paper award winners in engineering," *English Specific Purpose*, vol. 32, no. 3, pp. 157–169, 2013. [Online]. Available: https://doi.org/10.1016/j.esp.2013.04.002
[34] J. E. Harmon, "Current contents of theoretical scientific papers," *J. Tech. Writing Commun.*, vol. 22, no. 4, pp. 357–375, 1992. [Online]. Available: https://doi.org/10.2190/v051-8uka-w8fj-u54n.
[35] M. Kuteeva and L. McGrath, "The theoretical research article as a reflection of disciplinary practices: The case of pure mathematics," *Appl. Linguist.*, vol. 36, no. 2, pp. 215–235, 2015. [Online]. Available: https://doi.org/10.1093/applin/amt042
[36] L. L. Khaw and W. W. Tan, "Creating contexts in engineering research writing using a problem-solution-based writing model: Experience of Ph. D. students," *IEEE Trans. Prof. Commun.*, vol. 63, no. 2, pp. 155–171, Jun. 2020.
[37] H. Silyn-Roberts, *Writing for Science and Engineering: Papers, Presentations and Reports*. Amsterdam, The Netherlands: Elsevier, 2013.
[38] P. A. Laplante, *Technical Writing: A Practical Guide for Engineers, Scientists, and Nontechnical Professionals*. Boca Raton, FL, USA: CRC, 2018.
[39] J. Zobel, *Writing for Computer Science*, 3rd ed. London, UK: Springer, 2014.

Gerald Rau received the Ph.D. degree in Plant Breeding from Cornell University, Ithaca, NY, USA, in 1989. He teaches with the Electrical Engineering Department, National Chung Cheng University, Chiayi, Taiwan. His current research interest focuses on engineering writing.

Teaching Case

Designing STEM-Specific Student-Friendly Reading Content for the Engineering English Classroom

—Divya John and G. Sandhiya Devi

***Abstract*—Introduction:** *Teachers of English for specific purposes (ESP) face challenges when helping engineering students to comprehend discipline-specific reading materials because these students have not been exposed to engineering reading materials at the secondary-school level. This study examines how to best create Science, Technology, Engineering and Mathematics (STEM)-specific, student-friendly reading materials for engineering learners to make the transition from general topics to engineering topics comfortable.* **About the case:** *Working at affiliated colleges of Anna University, Chennai, India, we substituted reading materials for those used in textbooks to improve our students' receptivity to reading classes. We discuss reading as a skill, the level of reading comprehension needed in the engineering context, and its relevance to technical and professional communication.* **Situating the case:** *The study focuses on the need to design materials for the engineering classroom.* **Methods:** *Our methods include a present-situation analysis to examine the challenges that teachers encounter, and analysis of qualitative feedback on the reading materials and activities introduced in the classroom.* **Results and discussion:** *The paper includes our experiences in designing reading materials: conducting the present-situation analysis, finding and designing appropriate reading materials, creating and implementing reading activities, and collecting feedback. The responses of the learners indicate that introducing STEM-specific material may increase the learners' participation and improve their comprehension skills.* **Conclusion:** *The study shows that STEM-specific, student-friendly reading materials fostered a positive attitude and improved, high-level comprehension.*

Index Terms*—Reading comprehension, reading materials in English for specific purposes (ESP), reading skills, STEM-specific reading materials.*

This teaching case focuses on introducing student-friendly reading materials specific to science, technology, engineering, and mathematics (STEM) fields to engineering students in the English classroom. We explore the pedagogical approach used in selecting reading materials from authentic and accessible resources, and designing activities based on them to enhance learner receptivity. We also focus on the implications for the teaching community.

Communication is one of the skills needed by engineers according to the ABET criteria, so courses designed for engineering students should meet their needs [1], [2]. Industry is aware that new engineering graduates joining the workplace often lack the necessary writing skills, are unable to make effective presentations, and lack both team and interpersonal communication skills, thus, posing a challenge to the companies that hire them. As a result, recent developments in engineering practice and globally accepted accreditation standards demand a transformation in education—a move that integrates technical and professional communication [3]. No one can deny the importance of reading when it comes to preparing students for effective communication. Revising and re-envisioning the reading-skills components of communication courses have become the need of the hour because existing textbooks fail to provide appropriate resources [4].

The study reported here was conducted in an engineering college affiliated with Anna University, a technological university in Chennai, India, where most of the students learned English as a second language in secondary school. However, the students' language skills are at different levels, ranging from A1 (most basic user) to C2 (most competent user) on the Common European Framework of Reference for Languages (CEFR). The course and the syllabus are designed by the university's curriculum designers, keeping in mind

Manuscript received August 14, 2020; revised March 26, 2021; accepted March 26, 2021. Date of publication October 20, 2021; date of current version November 24, 2021. *(Corresponding author: Divya John.)*
Divya John is with the Department of English, Sri Sivasubramaniya Nadar College of Engineering (SSN College of Engineering), Chennai, Tamil Nadu 603110, India (email: divyajohn@ssn.edu.in).
G. Sandhiya Devi is with the Department of English, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu 600124, India (email: sandhiyadevi.g@ritchennai.edu.in).

IEEE 10.1109/TPC.2021.3110419

Practitioner Takeaway

- Engineers need to develop vital reading and reading comprehension skills during their undergraduate days because those skills enhance technical and professional communication.
- ESP teachers need to create student-friendly reading materials to prepare students to read and understand complex texts.
- To design reading materials, instructors should conduct a present situation analysis, find and design reading materials, create and implement reading activities, and collect feedback.

the multilevel nature of our classrooms. The first-year engineering program offers two courses, Communicative English and Technical English, to help students attain the language skills that they will need for the engineering program. The syllabi of both courses follow an integrated-skills pattern based on listening, speaking, reading, and writing (LSRW). This case examines the reading component of these syllabi, where teachers choose their reading materials either from the prescribed textbooks, the internet, or elsewhere, based on the students' comprehension level in English.

This teaching case poses two research questions.

RQ1. What are the teachers' constraints and experiences in the process of designing STEM-specific, student-friendly reading materials in the ESP classroom?

RQ2. What are the students' views of STEM-specific, student-friendly reading materials?

Our two-week study was conducted to give the learners sufficient input on STEM-related topics, thereby enhancing their range of comprehension from the existing level to a more advanced level. STEM-specific topics helped the students involve themselves in and enjoy the activity, and helped them develop a positive attitude toward reading in their discipline.

About the Case

English is the medium of instruction in the majority of the tertiary-level institutions in India. Most of the engineering programs offered by Anna University include courses in English in the first year to bridge the gap caused when students move from the general topics that they encountered in secondary school reading to discipline-specific topics in their engineering courses. The study was prompted by the engineering learners' lack of proficiency in reading skills needed in their engineering courses.

Language teachers face several challenges when teaching reading in engineering colleges because the textbook reading selections are outdated and do not engage the students' interest. Though the course syllabus gives equal weight to all the LSRW skills, teachers focus on grammar and writing, and do not spend enough time concentrating on reading skills in the classroom because they assume that students are able to read at the necessary level of proficiency and understand what they read. As far as the learners' approach to reading is concerned, their lack of sufficiently advanced reading skills prevents them from achieving the needed score in reading proficiency examinations. Because of this problem, engineering learners need to be introduced to STEM-specific reading materials to develop familiarity with engineering topics and the needed level of reading expertise in English.

Research on enhancing reading skills has been pursued by experts in English language teaching (ELT), applied linguistics, second language acquisition, and English for specific purposes (ESP). ESP teachers, who design reading materials, must keep in mind the bottom-up and top-down approaches. When readers create meaning from letters, words, phrases, clauses, and sentences by sorting out a text into phonemic units that signify lexical meaning, and then construct meaning in a linear manner, they are using a bottom-up approach. According to this approach, the reading task can be understood by considering it as a sequence of steps in a fixed order—from sensory input to comprehension and appropriate response [5, p. 33]. When readers use their knowledge of syntax and semantics to be independent of the print and phonics of the text, they are using a top–down approach. Hudson identifies four processes: predicting, sampling, confirming, and correcting [5, p. 37]. Goodman called the top–down approach a psycholinguistic guessing game [5].

Using this approach, readers not only guess the meaning of the text but also use their internal concept of language, past experiences, and conceptual background to generate the meaning of the text [5, p. 37]. Hudson believes that a successful reader must overcome the issues in reading by interacting with the text [5]. The increased familiarity helps the reader to gain letter recognition and higher level cognitive skills. According to Grabe, the interactive process is the outcome of the reader's interaction with the text along with background knowledge at the time of comprehension [5, p. 39].

The schema theoretical view of reading indicates that the knowledge of relevant schema, whether content schema or formal schema, is essential to read any kind of text. When the schema is activated, the reader's previous knowledge interacts with the new information gained while comprehending the text. Thus, interpreting the text becomes easier [5, p. 260]. Alderson states that the mental schema of students can be activated by making them read text that they are familiar with [6]. To activate the mental schema while reading, suitable texts should be selected and the right task assigned. The lack of proper schematic knowledge results in difficulty in understanding the text. Thus, when learning to read, schema activation plays a key role in understanding the text.

According to Anderson, the explicit teaching of reading strategies in the classroom is very effective in improving reading comprehension [7]. Strategies are not isolated actions but rather the result of a process of planning more than one action to accomplish a second-language task [7, p. 10]. Grabe holds that skills are automatic, whereas strategies are intentional and deliberate [8]. Metacognitive strategies are used especially by gifted readers [5, p. 292]. These skills help learners separate important information from the less important, and to track their attention while reading the text [5, p. 113]. Newton et al. have observed that academic reading tasks entail reading to find answers; learn new ideas and related details; develop expertise; synthesize information from long or multiple texts; look for evidence, arguments, or positions; or critique someone else's ideas [9].

In an engineering program, the minimum requirement for an ESP course is developing reading comprehension abilities. So the students in the engineering program have to complete the reading lists assigned to them in their course. According to Jordan, the main strategies, skills, and subskills utilized for reading are prediction; skimming; scanning; distinguishing between factual and nonfactual, important and less important, relevant and irrelevant information, explicit and implicit information, ideas, examples, and opinion; drawing inferences and conclusions; deducing unknown words; and understanding graphic presentations, text organization, and the linguistic/semantic aspects of the passages [10].

Dudley-Evans and St. John state that a significant contribution to the teaching of reading in ESP courses is the balance between skills and language development. Some key reading skills can be learned and transferred into the new language: selecting what is relevant for the current purpose; using the features of the text such as heading, layout, typefaces; skimming for content and meaning; scanning for specifics; identifying organizational patterns; understanding relations within a sentence and between sentences; using cohesive and discourse markers; predicting, interfering, and guessing; identifying main ideas, supporting ideas, and examples; processing and evaluating the information during reading; and transferring or using the information while or after reading [11, pp. 96–98].

Subsequent to reviewing related ESP literature and ELT specialists, Pritchard and Nasr furnish a reading comprehension checklist: understanding the gist of a text (skimming); locating specific information (scanning); understanding explicit and implicit information stated in a text, information from figures, diagrams, and tables, imperative and instructional language, and referents; recognizing synonyms in similar contexts; inferring meaning by prefixes, suffixes, and word families; recognizing nominal compounds; and summarizing and drawing conclusions about a text [12].

The reading portion of the present Anna University syllabus entails most of these skills and subskills of reading, leaving the choice of the text to inculcate these skills to the teacher. The course objective in reading is to help first-year engineering students' transition from general interest topics to technical and engineering topics. We attempted to discover whether introducing STEM-related topics to engineering students would help them better comprehend the engineering topics and how receptive they were to these topics. The study suggests simple strategies in designing student-friendly materials that teachers can use to make their reading classes more dynamic.

Wolfe's analysis of technical communication texts reveals that these textbooks disappoint engineering students and requests that teachers re-envision technical communication courses [4]. Paretti et al. show that an integrated approach to learning technical content and professional communication skills leads to the students' holistic development as engineers, and thus enhances their capacity for transition to engineering workplaces [1]. Hodges and Seawright discuss the challenges that instructors encounter in preparing students for professional communication in educational settings; they hint at adapting and revising the curriculum to suit transnational workplaces [13].

Engineering communication involves publicizing research and marketing products to masses. Publicizing research can be done by writing about it in technical journals and popularizing it through people-friendly magazines. To market products through social media like newspapers or magazines, engineers need to use simple language. Whether spoken or written communication, it is better to communicate in a plain language style that reaches a larger audience [14]. Hence, in this study, three STEM-related magazines were chosen to assist students in speaking and writing in a simple style to suit the needs of general listeners and readers.

Boiarsky discusses introducing mini-lessons on reading as a part of a sequence of technical communication assignments. She adds that teaching engineering students how to learn provides them with skills and strategies necessary to tackle new assignments in their jobs [15]. Artemeva et al. designed a discipline-specific communication course using theories of genre and situated learning, and argued that a communication course can be tailored to any discipline provided three conditions are met: assignments must be connected to the subject matter, a dialogic environment must be planned, and the nature of assignments must allow students to build on their learning experience [16]. In response to the calls of MIT alumni for better instruction in communication, Craig et al. give three case-studies that introduce students to target competencies of professional practice, effective teamwork and collaboration, and understanding and arguing visual data in the communication-intensive curriculum [17].

Ford argues that students appeared to transfer rhetorical strategies like audience awareness, sense of purpose, organization, use of visuals, professional appearance, and style from the technical communication classroom to the engineering classroom and to the workplace, a process that, in turn, helped them to be better communicators [18]. Paretti and McNair state that engineering curricula confront challenges in communication. For example, engineering faculty lack expertise in communication skills, and technical communication specialists lack content knowledge about engineering as well as current engineering practices in the workplace [19]. The nuances of communication cannot be learned in a single writing course. With the tag line "beyond writing-intensive (WI) to department-wide," Patton argues rightly for integrating writing throughout the engineering curriculum, and presents a case-study showing how an integrated communication curriculum can be implemented [20].

Situating the Case

Designing reading materials for the engineering classroom is demanding because of the processes involved and the research required, but unfortunately, not many teachers are aware of the studies that have been done on these topics. Although many professional communication studies address spoken or written communication, research on reading is scant, despite the fact that proficient reading skills are vital for engineers.

Nevertheless, several scholars have examined the design of material in the ESP and English for academic purposes (EAP) contexts. Hutchinson and Waters assert that an ESP teacher has to deal with needs analysis, syllabus design, and material-writing or adaptation and evaluation [21]. ESP teachers need not learn specialist subject knowledge but need only three things: a positive attitude toward ESP content, the knowledge of the fundamental principles of the subject area, and an awareness of how much audience members probably know already—that is, the ability to ask intelligent questions [21, p. 163]. Jordan suggests that the teachers of ESP courses need to pose questions to class members, team teach with subject matter experts, address discipline-specific topics and texts, and utilize individualization and self-access [10] in designing assignments.

Spector-Cohen et al. provide a four-pronged approach in designing EAP reading courses at the university level: incorporating specific instruction on linguistic forms, using reading comprehension strategies, using examples of various academic

genres, and utilizing criterion tasks in assignments that focus on meaning and reflect real-world academic tasks [22]. They suggest three levels of guidelines.

1. At the beginning level, they suggest enhancing student proficiency using general and nonspecialized texts.
2. At the intermediate level, they suggest using examples that move toward authentic academic texts in various content areas and varied subject matter.
3. At the advanced level, they emphasize that texts should be content-specific and taken from a single academic discipline [22].

The students in our study were introduced to reading general texts in the first semester, and then STEM-specific materials in the second semester, before being exposed to discipline-specific reading materials in the later semesters.

According to Hyland, the context-specific nature of EAP is that no single textbook is ever ideal for a particular class [23]. Our study incorporates Hyland's view of creating material that will suit a specific class. Hyland demonstrates three main ways to achieve the validity and reliability of needs analysis: triangulation, prolonged engagement, and participant verification [23]. EAP teachers must emphasize the skills that are suitable and serve the purpose of the academic and professional communities. The use of authentic material for an EAP course reduces learning difficulties and creates confidence [24].

Pritchard and Nasr suggest selecting reading materials based on needs analysis; they conducted a reading improvement course using authentic engineering texts as opposed to simplified ones that enhanced student reading comprehension [12]. And Basturkmen proposes a three-level framework for ESP curriculum development: analyzing needs, investigating specialist discourse, and determining curriculum [25]. To design reading materials in an EAP setting, Kuzborska points to the teacher's constraints and opportunities during the decision-making process. She admits that in conducting needs analysis, formulating goals and objectives, finding material, creating activities, and using materials in class, the selection of materials was more led by intuition than based on research [26].

Referring to studies by Jordan and McDonough, Hirvela says that though students consider reading as the least difficult of skills, it is by far the most significant [27]. He describes the shift from register analysis to discourse analysis, rhetorical approach, and genre analysis [27]. Hirvela stresses the importance of introducing genre-based text during the reading course because genre-based instruction transfers "specialized literary reading" to "science-based writings" because genre-based reading instruction helps learners recall their prior knowledge that is reflected in their writing [27, pp. 87–89]. He states that the input-based and output-based approaches are the two directions by which the learners enhance their writing skills through the act of reading [28]. Thus, the transformational activity of "Reading for Writing" is achieved. In the input-based approach, learners use reading as input for writing; in the output-based approach, the focus is on the act of writing in which receptive skills are transferred to the productive skills of writing [28, p. 128].

Parkinson recommends that teachers of "English for Science and Technology" collaborate with the engineering discourse community to be better exposed to the students' discipline [29]. Thus, student needs can be fulfilled by a process of integration [29, p. 158]. Familiarizing the students with the target discipline creates a positive effect, improves their perceptions, and serves their needs in subject-specific contexts [29, p. 168]. Similarly, Woodrow asserts that to support learners, an ESP course should be designed based on their expectations, and the materials should have an academic or professional focus [30]. According to Anthony, to create an effective ESP course, the instructor must be aware of four pillars: learning objectives, materials and methods, need analysis, and evaluation [31]. Arnó-Macià et al. observe that the existing ESP courses should be adapted to the ever-changing needs of engineering undergraduates in a globalized world. So ESP course designers need to re-evaluate their courses to assist learners in both academic and specialized professional communication [32].

Methods

This teaching case followed a practitioner research paradigm. We first examined the process of designing reading materials for the engineering classroom taking into consideration the constraints and complexities experienced by teachers in designing student-friendly reading materials. Reflecting on the previous research conducted in the ESP and EAP contexts, we delved into the processes involved in designing reading materials: conducting a present-situation analysis (PSA), choosing suitable material, designing the reading

material and activities, testing the material in the class, and collecting qualitative feedback.

Citing Long and Berwick, Flowerdew suggests that both inductive and deductive procedures be used in needs analysis, so many elements must be considered [33]. The inductive procedure includes expert intuitions, participants' and nonparticipants' observations, and unstructured interviews. The deductive process uses surveys and questionnaires, structured interviews, and sometimes criterion-referenced performance tests. The tools used for our study were an informal need analysis, informal interviews, and classroom discussions. The informal needs analysis was intended to discover what kind of reading texts and topics the students preferred because analyzing learner needs is a part of designing materials for a course [21], [23], [25], [31], [34]. Subsequently, a PSA was performed from the point of view of both teachers and students before the actual intervention took place. The teachers' constraints and insights were also noted as part of PSA. As far as the issues of students are concerned, the PSA dealt with the learners' lacks and wants, and the target-situation analysis considered the learners' needs [33].

The framework of the study was two weeks, with four 1-hour periods each week. The participants were first-year engineering students 18–19 years of age, from the authors' colleges. In all, 53 students from computer science engineering, 64 from electronics and communication engineering, and 61 from chemical engineering participated. The English skills of these second language learners ranged from A1 (basic users) to C1 (proficient users) on the CEFR scale. The students gave their informed consent to participate in the intervention. They were introduced to some texts and activities from *National Geographic*, *Discover*, and *IEEE Spectrum*. After the intervention, detailed written feedback was collected from them in which they recorded their experience in reading STEM-specific reading material; the feedback was coded thematically.

Results and Discussion

This section deals with the procedures that we followed in designing and using STEM-specific, student-friendly reading material, and the constraints and insights that we experienced. The processes included conducting the PSA, choosing suitable material, designing the reading material and activities, testing the material in the classroom, and collecting qualitative feedback.

Conducting the PSA The courses on which our study was based had presented several problems in previous iterations because the reading topics in the prescribed textbooks were not up-to-date. By the time books are published, the topics in them are frequently not as appealing to the students as articles in current STEM-related magazines. Also, though the prescribed syllabus expects that all of the LSRW skills will be covered, not all of them receive sufficient attention partly as the result of time constraints. Thus, attention to reading skills frequently suffers due to a lack of sufficient time. Moreover, students resist reading in the class because they do not want to read anything new, and also because of other predictable reasons. Therefore, teachers tend to concentrate on topics that students are not familiar with. In addition, teachers assume that students know how to read and comprehend a text, and think it is better to skip the reading part of the course in the blind hope that they have acquired the strategies of reading and reading comprehension needed to pass university examinations. We conducted an informal needs analysis to find out what kind of reading texts our students preferred by questioning them on their needs. This was done through informal interviews and classroom discussions. In general, students demanded interesting science-related materials.

It is not always feasible to conduct a needs analysis on an elaborate scale. Preparing the questionnaire, administering it to the students, getting their feedback, analyzing the results, and arriving at a conclusion is time-consuming. Experts assert that to guarantee student proficiency in ESP reading, the English teacher must create materials that best meet the students' needs [21], [25]. Anthony confirms that the ESP instructor should introduce materials that serve the learners' needs in their respective discipline [31]. Thus, we decided to design STEM-specific reading materials for engineering learners to provide authentic reading texts that would enable them to become familiar with STEM-related topics at a level that they were comfortable with as beginning engineering students.

Choosing Suitable Materials A teacher has to be cautious when choosing reading materials for the engineering classroom because recent studies reveal that the prescribed textbooks may not provide learners with the background required in today's workplace. Learners at the tertiary level are expected to develop their academic and professional competence ranging from academic reading to advanced level language reading. In choosing a

text, an ESP teacher has to balance between needs and motivational factors by offering good materials, selecting materials appropriate to the students and their areas of study, being creative with what is available, modifying activities to suit the learners' needs, and providing supplementary activities and input [11, p. 173]. In our study, we chose the materials from three sources.

1. *National Geographic* (https://www.nationalgeographic.com/magazine/), the official magazine of the National Geographical Society, which publishes on topics in science, geography, history, and world culture
2. *Discover* (https://www.discovermagazine.com/magazine), a magazine that reports the latest developments in science, medicine, technology, and the environment
3. *IEEE Spectrum* (https://spectrum.ieee.org/), the flagship magazine of IEEE, that informs its readers about major trends and developments in technology, engineering and science

These three sources serve as authentic texts.

The students will be strongly motivated to learn a language if they find a close connection between the content of the material and their study needs, and such material is most likely to be subject-specific [10, p. 262]. When a topic is selected for ESP use, it should be subject-specific, introducing the learners to the reading genres and discourses [23]. Grellet believes that a text with a focus on "global understanding" creates a deeper understanding among learners, and their confidence is strengthened when they are exposed to authentic texts with appropriately difficult vocabulary [35, p. 6]. Stoller says that in EAP contexts, materials and tasks serve as vehicles through which students are prepared to focus on future academic pursuits; therefore, EAP teachers should utilize materials like magazines, newspapers, and other science articles to meet the existing demands [36, pp. 577–579]. It is also understood that EAP materials and tasks should prepare students to handle even complicated texts and tasks individually [36, p. 580]. Specialists in the field affirm that text authenticity is very important in content-specific EAP courses [23], [30], [37].

Some of the topics dealt with in our classes were chosen by the engineering students themselves when they were exposed to these magazines during laboratory classes. First, teachers should be given autonomy to design and implement new materials that will help students become "independent" and "self-sufficient" readers. Second, learners should be given autonomy to select their own topics as provided in our study. The teacher-generated materials and learner-generated materials help eliminate monotony in the class [11].

Designing the Reading Material and Activities

According to Hutchinson and Waters, when selecting a text, the instructor should choose a naturally occurring piece of communication suited to the learners' needs and interests, and capable of generating useful classroom activities [21]. In the present study, reading materials were chosen from articles in *National Geographic*, *Discover*, and *IEEE Spectrum*, and reading activities were designed based on these articles. (A broad outline of the activities and skills is given in Table I.)

As suggested by experts in the subject, introducing these reading materials was intended to help students achieve the necessary reading competency required at the university level, and the activities were designed based on the skills and subskills of reading for the same reason [10]–[12]. The activities were designed to test the students' reading comprehension with multiple choice and true or false questions, fill in the blank statements, and logical sequences of sentences. The tasks assigned had various levels of difficulty, and classroom discussion was encouraged [35, p. 11]. The learners in this study were also involved in pair work, group work, presentations, and discussions. Having an ESP expert take the role of material designer is a common practice in the ESP scenario [11], [21]. Hutchinson and Waters propose a model of four elements in designing ESP course materials: input, content focus, language focus, and task [21]. Watkins suggests numerous activity types that can be adapted in the ESP classroom [38].

Choosing reading materials is challenging for teachers, as they have to read several passages before deciding on which passages to select. On one hand, a lack of in-depth content knowledge can be a constraint for the teachers. On the other hand, it is easy for the students to criticize selections as "boring," "outdated," and the like, thus, negating by a single stroke all the teacher's efforts.

Designing activities can be very strenuous and time consuming for teachers who are busy with a tight schedule. If they want to design activities by expending less effort, then they can choose group discussion as it is an activity that evaluates

TABLE I
READING ASSIGNMENTS, SKILLS TAUGHT, AND ACTIVITIES

Magazine	Assigned Article or Topic Selection Method	Skills Taught	Activities
National Geographic Magazine	"Extreme black hole vindicates Einstein"	• Predicting • Skimming • Scanning for specifics	• Comprehension questions • True or false • Unscrambling and matching words
	"Epic engineering rescued colossal ancient Egyptian temples from floodwaters"	Identifying main ideas and supporting ideas	• Comprehension questions • Fill in the blanks • Discussion: Contributions and steps taken by UNESCO to save monuments
	"Why does earth have a moon and how does it affect our planet?"	Skimming to obtain the general impression of the text	• Reading comprehension questions • Fill in the blanks
	"This is now the world's largest volcano, geologists say— Mountain Tamu Massif"	Activating metacognitive skills to build logical reasoning	• Reading comprehension questions • Rearranging jumbled sentences • Unscrambling words
	"Coronavirus spikes outside China show travel bans aren't working"	Applying metacognitive skills to analyze problems	• Group Discussion
Discover Magazine	"How do autonomous cars work?"	• Understanding the text, • Opinions / facts	• Discussion • Summarizing discussions
	"What science says about why you're stressed and how to cope"	• Interpreting images, • Drawing conclusions about the text	• Group Discussion • Summarizing discussions
	Student groups choose their own topics.	• Summarizing information, • Presentation skills	Presenting topics
	Student-groups choose a month.	Synthesizing information from larger texts	Presenting science news
	Students choose their own topics.	Understanding the gist of the text, Intensive reading	Writing summary
IEEE Spectrum	Student groups are assigned topics from the magazine and select their own subtopics.	Processing and evaluating information from the text	Summarizing and presenting the research article
	Students choose articles of interest.	Summarizing	Practice in presentation skills

students more accurately than multiple choice questions. During a group discussion activity, the teacher can also assess clearly a student's level of comprehension of the material within a short time.

Testing the Material in the Classroom
Hutchinson and Waters provide an exhaustive checklist for material evaluation in an ESP course because it is essential to check the fitness of the material for a particular purpose [21]. Jordan says that once the material is selected, it can be judged only after implementing it successfully in the classroom, and then collecting feedback [10]. Piloting and evaluating the new materials by both the learners and the teachers in the classroom are essential [37].

In our study, after designing the reading material, we tested it in the classroom. During the activities, we observed that the learners were highly motivated by the texts as well as their content. Because all the learners were involved in the study, the group work, pair work, and discussion promoted a student-friendly learning atmosphere in the classroom. The passages chosen for reading practice gratified the average learners' curiosity and encouraged them to perform better. The students had extensive exposure to new words and gained confidence to read longer texts without much difficulty. The students enjoyed group work and group discussion more than multiple choice questions, fill in the blanks, and true–false responses. Constant practice and close observation by the teacher helped them to develop their comprehension ability. If designing activities for the class is time consuming, it is good to have students generate a topic and conduct discussions based on it.

Collecting Qualitative Feedback Experts believe that students' views and perceptions should be considered while analyzing and designing courses [39]. Accordingly, after completing our study, we collected and thematically coded the students' views. Some of their observations are given below.

Several students confirmed that the reading materials and subsequent activities were useful and informative. They suggested that more of such activities should be conducted frequently.

> We read many interesting accounts and learned several facts.

> Reading comprehension classes are different from the regular classes. They save time and are very informative.

> Doing the activity was a pleasant experience. The passages chosen for the activities were interesting.

> Reading the articles was a good experience. I learned a lot of noteworthy facts ….

Some students said that to find the answers, they applied a few reading strategies that worked out successfully. They also used reading strategies, such as skimming and scanning: skimming to get the basic gist of the text and to promote quick reading, and scanning to reject less important or irrelevant information found in the text. The feedback indicated that the students who employed skimming and scanning were able to identify the question part (agreeing with the information in the text) whether the information is "Given" or "Not Given" in the passage.

> Reading magazine articles was very useful and productive. It was extremely encouraging. We were eager to do more such activities. This helped me to understand the strategies in reading.

> This activity is a good reading practice. It would be very useful during the exam, and also helpful to take entrance exam in future confidently.

> The reading comprehension activity helps us to find the answers within a short time.

> It is really very useful. I learned many words. It was interesting to guess their meaning.

> I made use of skimming and scanning strategies for the answers. We had fun doing these activities.

The students commented that though the activities were conducted for only a short period, they were grateful for the new words they learned and the unique experience they gained.

> Reading comprehension activities were extremely useful. Our vocabulary and reading skills improved.

> The activity is not only interesting, but it also improves our vocabulary and enables us to gain knowledge.

> The reading activities were actually learning lessons—my vocabulary developed.

> The activities provided a kind of mental refreshment.

Some students expressed apprehension when they encountered new words. A few conveyed their concern about dealing with engineering topics that

they considered boring. They felt such fact-related topics should be avoided in an English class, which can be made interesting with short stories and other forms of literature. On the whole, however, the reading comprehension materials kept most of the students active in the classroom.

> Learning new topics is quite interesting and refreshing. It breaks the monotony of the other subject classes.
>
> The activity was conducted in an interesting manner. So we were attentive and our comprehension level increased.
>
> A few students claimed to have developed their reading, thinking, and writing skills.
>
> The activities developed our reading and writing skills.
>
> This reading comprehension practice is really enlightening in the development of language. By doing this, my reading skill improved.
>
> I think my reading skills and thinking skills have improved after reading the magazine articles.
>
> It was an excellent experience. I look forward to doing these assignments as they are informative.
>
> The students also valued the group discussions.
>
> Reading comprehension activities were extremely enriching. We were able to discuss the answers in the class during the group discussions.
>
> Group discussions cultivated our argumentative skills.
>
> … my language skills developed during the group discussions.

Although our study occupied only two weeks of a semester-long course, the students realized the importance of taking a practice test on reading to overcome their reading issues. The activity not only promoted discussions among the students but also motivated them to read the recently published magazine articles. Thus, their inhibitions about reading comprehension were reduced.

Conclusion

This study examined reading and reading comprehension as skills that ESP learners need to develop. It also addressed the need for ESP practitioners to choose reading materials and design activities based on them for the learners. We have reviewed the process involved in creating STEM-specific, student-friendly reading materials for engineering learners in the ESP classroom. We have discussed the constraints involved in conducting a PSA, finding and designing reading material, creating reading activities and implementing them, and collecting student feedback. The learners' responses indicate that introducing STEM-specific, student-friendly material increased their participation.

When designing activities based on the reading materials, many teachers resort to the materials and the activities in the textbook, which are handy. They face several constraints in the design process: conducting the PSA, which may be rather difficult at times; organizing an elaborate needs analysis; and choosing interesting and student-friendly reading materials, which can be challenging. Other major concerns are that teachers may be hindered by busy schedules, they may lack knowledge in the content area, and they may not have done much research in ESP and engineering communication.

Implications We have reviewed in detail the processes involved in creating STEM-specific, student-friendly reading materials for engineering learners in the ESP classroom. When designing activities based on the reading materials, many teachers simply use the text passages and activities in the textbook. In fact, teachers who want to design reading materials for their students can examine the constraints and guard against them, and implement the necessary solutions by conducting a PSA, finding and designing the reading materials, creating and implementing reading activities, and finally collecting feedback. Doing so will increase their own knowledge and that of their students.

Paretti et al. focus on assisting students to learn to think and write like globally competent engineers because contemporary engineers must read and understand complex texts [3]. It is the duty of engineering communication teachers to provide their students with plenty of reading materials that are interesting, varied, attractive, abundant, and accessible as Grabe and Stoller suggest [40]. By reading such texts representing multiple genres, students learn specific vocabulary and enhance their language skills. When teachers face time constraints in preparing and designing activities, they can simply assign a reading passage and then assign a discussion based on it. Another option is using learner-generated topics, followed by discussions and presentations based on them [11]. Teachers should share their experiences designing and implementing activities at seminars and

conferences, and in journal articles to help them develop as researchers.

Limitations The present study was done on a small scale over only two weeks. The same study should be conducted for a longer duration and the impact assessed for value-added results. In addition, if the study had involved more classes and mixed-ability groups, it would have yielded more compelling results. This study involved the work of only two teachers during the designing and testing period. Conversely, if the views of a large number of teachers were collected, the study would have been richer.

Future Research Reading skill is almost entirely neglected in the field of engineering education. As a result, engineers often lack the skill to read and understand advanced texts. The power of reading skills cannot be denied. The integrated syllabus of 60 1-hour periods of LSRW skills at our university provides only 15 periods for the reading component. In fact, if an engineering curriculum was constructed with due importance given to reading, better engineers who think critically and communicate effectively would be produced. Likewise, if student-friendly reading materials were designed, tested, and published, they would be beneficial to the teaching community and consequently to the student community.

This study was related to the process of designing reading materials for beginning engineering students. Quantifying the results of using such materials could measure their impact on students. In the same way, duplicating our study across a large number of institutions would produce more engineers with excellent reading comprehension and communication skills.

References

[1] M. C. Paretti, A. Eriksson, and M. Gustafsson, "Faculty and student perceptions of the impacts of communication in the disciplines (CID) on students' development as engineers," *IEEE Trans. Prof. Commun.*, vol. 62, no. 1, pp. 27–42, Mar. 2019.
[2] N. T. Buswell, B. K. Jesiek, C. D. Troy, R. R. Essig, and J. Boyd, "Engineering instructors on writing: Perceptions, practices, and needs," *IEEE Trans. Prof. Commun.*, vol. 62, no. 1, pp. 55–74, Mar. 2019.
[3] M. C. Paretti, L. D. McNair, and J. A. Leydens, "Engineering communication," in *Cambridge Handbook of Engineering Education Research*, A. Johri and B. M. Olds, Eds., Cambridge, UK: Cambridge Univ. Press, 2014, pp. 601–632.
[4] J. Wolfe, "How technical communication textbooks fail engineering students," *Tech. Commun. Quart.*, vol. 18, no. 4, pp. 351–375, Sep. 2009.
[5] T. Hudson, *Teaching Second Language Reading*. New York, USA: Oxford Univ. Press, 2007.
[6] J. C. Alderson, *Assessing Reading*. Cambridge, UK: Cambridge Univ. Press, 2000.
[7] N. J. Anderson, *Practical English Language Teaching Reading*. New York, NY, USA: McGraw-Hill, 2008.
[8] W. Grabe, *Reading in a second language: Moving from Theory To Practice*. Cambridge, UK: Cambridge Univ. Press, 2009.
[9] J. M. Newton, D. R. Ferris, C. C. Goh, W. Grabe, F. L. Stoller, and L. Vandergrift, *Teaching English to Second Language Learners in Academic Contexts: Reading, Writing, Listening, and Speaking*. New York, NY, USA: Routledge, 2018.
[10] R. R. Jordan, *English for Academic Purposes: A Guide and Resource Book for Teachers*. Cambridge, UK: Cambridge Univ. Press, 1997.
[11] T. Dudley-Evans and M. J. St John, *Developments in English for Specific Purposes: A Multi-Disciplinary Approach*. Cambridge, UK: Cambridge Univ. Press, 1998.
[12] R. M. Pritchard and A. Nasr, "Improving reading performance among Egyptian engineering students: Principles and practice," *English Specific Purposes*, vol. 23, no. 4, pp. 425–445, Mar. 2004. [Online]. Available: https://doi.org/10.1016/j.esp.2004.01.002
[13] A. Hodges and L. Seawright, "Writing in transnational workplaces: Teaching strategies for multilingual engineers," *IEEE Trans. Prof. Commun.*, vol. 62, no. 3, pp. 298–309, Sep. 2019.
[14] K. S. Campbell, N. Amare, E. Kane, A. D. Manning, and J. S. Naidoo, "Plain-style preferences of US professionals," *IEEE Trans. Prof. Commun.*, vol. 60, no. 4, pp. 401–411, Dec. 2017.
[15] C. Boiarsky, "Teaching engineering students to communicate effectively: A metacognitive approach," *Int. J. Eng. Educ.*, vol. 20, no. 2, pp. 251–260, 2004.
[16] N. Artemeva, S. Logie, and J. St-Martin, "From page to stage: How theories of genre and situated learning help introduce engineering students to discipline-specific communication," *Tech. Commun. Quart.*, vol. 8, no. 3, pp. 301–316, 1999.
[17] J. L. Craig, N. Lerner, and M. Poe, "Innovation across the curriculum: Three case studies in teaching science and engineering communication," *IEEE Trans. Prof. Commun.*, vol. 51, no. 3, pp. 280–301, Sep. 2008.
[18] J. D. Ford, "Knowledge transfer across disciplines: Tracking rhetorical strategies from a technical communication classroom to an engineering classroom," *IEEE Trans. Prof. Commun.*, vol. 47, no. 4, pp. 301–315, Dec. 2004.

[19] M. C. Paretti and L. D. McNair, "Introduction to the special issue on communication in engineering curricula: Mapping the landscape," *IEEE Trans. Prof. Commun.*, vol. 51, no. 3, pp. 238–241, Sep. 2008.
[20] M. D. Patton, "Beyond WI: Building an integrated communication curriculum in one department of civil engineering," *IEEE Trans. Prof. Commun.*, vol. 51, no. 3, pp. 313–327, Sep. 2008.
[21] T. Hutchinson and A. Waters, *English for Specific Purposes*. Cambridge, UK: Cambridge Univ. Press, 1987.
[22] E. Cohen, M. Kirschner, and C. Wexler, "Designing EAP reading courses at the university level," *English Specific Purposes*, vol. 20, no. 4, pp. 367–386, Dec. 2001. [Online]. Available: https://doi.org/10.1016/S0889-4906(00)00019-3
[23] K. Hyland, *English for Academic Purposes: An Advanced Resource Book*. London, UK: Routledge, 2006.
[24] K. Hyland, "Specificity revisited: How far should we go now?," *English Specific Purposes*, vol. 21, no. 4, pp. 385–395, May 2002. [Online]. Available: https://doi.org/10.1016/S0889-4906(01)00028-X
[25] H. Basturkmen, *Developing Courses in English for Specific Purposes*. New York, NY, USA: Palgrave Macmillan, 2010.
[26] I. Kuzborska, "Teachers' decision-making processes when designing EAP reading materials in a Lithuanian University setting," *J. English Acad. Purposes*, vol. 10, no. 4, pp. 223–237, Dec. 2011. [Online]. Available: https://doi.org/10.1016/j.jeap.2011.07.003
[27] A. Hirvela, "ESP and reading," in *The Handbook of English for Specific Purposes*. B. Paltridge and S. Starfield Eds., West-Sussex, UK: Wiley-Blackwell, 2013, pp. 77–94.
[28] A. Hirvela, "Academic reading into writing," in *The Routledge Handbook of English for Academic Purposes*. K. Hyland and P. Shaw Eds., New York, NY, USA: Routledge, 2016, pp. 127–138.
[29] J. Parkinson, "English for Science and Technology," in *The Handbook of English for Specific Purposes*. B. Paltridge and S. Starfield Eds., West-Sussex, UK: Wiley, 2013, pp. 155–174.
[30] L. Woodrow, *Introducing Course Design in English for Specific Purposes*. New York, NY, USA: Routledge, 2017.
[31] L. Anthony, *Introducing English for Specific Purposes*. New York, NY, USA: Routledge, 2018.
[32] E. Arnó-Macià, M. Aguilar-Pérez, and D. Tatzl, "Engineering students' perceptions of the role of ESP courses in internationalized universities," *English Specific Purposes*, vol. 58, pp. 58–74, Apr. 2020. [Online]. Available: https://doi.org/10.1016/j.esp.2019.12.001
[33] L. Flowerdew, "Need analysis and curriculum development in ESP," in *The Handbook of English for Specific Purposes*, B. Paltridge and S. Starfield, Eds., West-Sussex, UK: Wiley, 2013, pp. 325–346.
[34] J. D. Brown, *Introducing Needs Analysis and English for Specific Purposes*. New York, NY, USA: Routledge, 2016.
[35] F. Grellet, *Developing Reading Skills Grellet: A Practical Guide to Reading Comprehension Exercises*. Cambridge, UK: Cambridge Univ. Press, 1981.
[36] F. L. Stoller, "EAP materials and tasks," in *The Routledge Handbook of English for Academic Purposes*, K. Hyland and P. Shaw, Eds., New York, NY, USA: Routledge, 2016, pp. 577–591.
[37] M. Charles, and D. Pecorari, *Introducing English for Academic Purposes*. New York, NY, USA: Routledge, 2015.
[38] P. Watkins, *Teaching and Developing Reading Skills: Cambridge Handbooks for Language Teachers*. Cambridge, UK: Cambridge Univ. Press, 2017.
[39] Z. M. Gaffas, "Students' perceptions of the impact of EGP and ESP courses on their English language development: Voices from Saudi Arabia," *J. English Acad. Purposes*, vol. 42, Nov. 2019, Art. no. 100797. [Online]. Available: https://doi.org/10.1016/j.jeap.2019.100797
[40] W. Grabe and F. L. Stoller, *Teaching and Researching Reading*. New York, NY, USA: Routledge, 2019.

Divya John received the Ph.D. degree in English Language and literature from the University of Calicut, Kerala, India, in 2008. She is currently an Associate Professor with the Department of English, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India. She is also an experienced teacher. She has authored or coauthored papers at 30 national and international conferences on ELT and many articles on ELT in national and international journals.

G. Sandhiya Devi received the M.Phil. degree in English from Madurai Kamaraj University, Tamil Nadu, India, in 2006. She is currently an Assistant Professor with the Department of English, Rajalakshmi Institute of Technology, Chennai, India. She is also an experienced teacher. She has authored or coauthored papers at four national and international conferences on ELT.

Teaching Case

Using Team-Based Learning to Promote Engineering Students' Performance and Self-Efficacy in a Technical Writing Class

—Shenghua Zha, Shenghua Wu, and Julie M. Estis

***Abstract*—Introduction:** *Technical writing is an essential skill set for engineering students. Many studies have been conducted, but very few have used experimental or quasiexperimental design to identify an optimal instructional method in a technical writing class.* **Situating the case:** *Team-based learning (TBL) is a well-structured learning method that prior studies have found to enhance students' academic performance. TBL includes individual and collaborative learning activities from lower to higher cognitive levels. Peer leadership, as evidenced in other studies, uses appointed student leaders to promote equal and active group participation and shows a potential to solve the gender issue found in engineering class collaborations.* **About the case:** *In this case, we infused peer leadership in TBL in three technical writing sessions of an engineering lab class. Appointed student leaders were responsible for initiating and sustaining discussions, asking each group member's input, and seeking collective decisions on solutions. The other class used traditional TBL activities.* **Methods/approach:** *Nonparametric analyses were conducted to compare students' technical writing skills and self-efficacy, as well as gender differences in two classes.* **Results/discussion:** *Students in the peer-led TBL class showed better technical writing skill retention than their counterparts in the traditional TBL class. The gender difference was identified in the traditional TBL class. However, we did not find any difference in students' self-efficacy between the peer-led and traditional TBL sections, though both observed a significant improvement at the end.* **Conclusions:** *We suggest studies with large sample sizes and equal distribution of female and male students.*

Index Terms*—Gender, peer leadership, team-based learning (TBL), technical writing.*

Technical writing has been identified as an essential skill set by educators and professional engineers. In total, 1500 engineers in Yalvac et al.'s [1] survey acknowledged that technical communication skills were critical in their career promotion and advancement. However, Gnanapragasam [2] pointed out that students chose engineering as their major because they were strong in math and science, not technical writing. Hence, it is crucial for engineering educators to find effective strategies to improve students' technical writing skills.

Many pedagogically focused studies have been conducted. Interdisciplinary collaboration is the most-discussed method [2], [3]. In Weissbach and Pflueger's [3] study, the course instructor collaborated with the university writing center. The instructor trained the writing tutors two weeks before tutoring for a specific assignment. The training aimed to remedy tutors' lack of disciplinary knowledge. Results showed that most students thought that the tutoring improved the quality of their reports. Other methods have also been found in publications on technical writing education, such as peer review [4] and project-based learning [5]. However, a recent review of publications on major technical writing journals, such as the IEEE Transactions on Professional Communication and the *Journal of Technical Writing and Communication*, showed that very few studies used an experimental or quasiexperimental research design [6].

The term "team-based learning" (TBL) has been used in many studies and includes many types of collaborative learning, such as collaborative exams or writing [7], [8]. We used TBL to refer to the well-structured collaborative learning method proposed by Michaelsen et al. [9]. In our teaching case, students worked in small teams and went through a series of individual and collaborative learning activities. At the beginning, readiness assurance tests and discussions targeted individual students' learning at the lower cognitive level. Questions at this stage (e.g., what is the best definition of literature review?) aimed to help

Manuscript received December 1, 2020; revised May 13, 2021; accepted May 14, 2021. Date of publication October 15, 2021; date of current version November 24, 2021. *(Corresponding author: Shenghua Zha.)*
The authors are with the University of South Alabama, Mobile, AL 36688 USA (email: shzha@southalabama.edu; shenghuawu@southalabama.edu; jestis@southalabama.edu).

IEEE 10.1109/TPC.2021.3110619

Practitioner Takeaway

- Many studies have been conducted to improve the technical-writing instruction at the organization, curriculum, and course levels. However, very few of them used the experimental or quasi-experimental design to identify an effective instructional method.
- The authors deployed the traditional (traditional TBL) and peer-led Team-based Learning (pTBL) methods in three technical writing sessions of two classes. They examined students' retained technical writing skills as well as their self-efficacy.
- Students' technical writing skills were significantly improved in two out of three sessions in the pTBL class. The gender difference was not observed in the pTBL class.
- Students in both traditional TBL and pTBL classes showed a significant improvement in their self-efficacy at the end of the study. No gender difference was observed in either class.

students strengthen their understanding and recall of concepts and principles. In the second half of the course, students solved application problems in teams. Those problems (e.g., draw a chart given the data in a short report) required them to generate, select, and compare the solutions.

TBL has been widely adopted in health-related education and has shown its effectiveness in improving students' accountability, collaboration skills, and academic performance [9]–[12]. Among those skills, accountability and collaboration are what the National Science Foundation and the engineering community have been promoting in the past 20 years [13]. Therefore, we believe that TBL deserves more attention than it currently receives in engineering and technical writing classes.

Gender inequality has been a big concern in collaborative learning in engineering classes [14]. Female representation in undergraduate engineering education has remained low, with a slight increase in the past decade [15]. As a result, male students have dominated most small-group learning implementations in engineering classes. Studies found that students' gendered behavior was magnified in one-gender-dominated small groups [16]–[20]. For example, male students used computers more and spent less time on writing than their female group members in group work. Those behaviors have resulted in a gendered difference in students' performance and self-efficacy. However, studies have not reached consistent conclusions regarding gender differences and their impact on technical writing classes.

Peer leadership may be a solution to the negative gendered impact. Although the literature shows that student leader selection methods and their responsibilities varied in different peer-led learning methods, the commonality was that student leaders were appointed by course instructors and trained before they started [21]–[23]. In addition, the training usually included group management and facilitation strategies, such as how to ensure equal participation. Studies have shown that peer leadership helped students fulfill collaboration expectations and improve performance [24], [25].

In the following section of this report, we describe how we implemented peer-led TBL (pTBL) in one section and traditional TBL in the other section of an undergraduate lab course (class). We then explain how this teaching case is situated in the literature of TBL, peer-led learning, and the impact of both on students' learning and gendered issues. Next, we depict the methods used to examine the impact of pTBL and TBL on students' technical writing skills and self-efficacy. After presenting the results, we discuss the practical and research implications of this case.

We considered the following research questions.

1. How do students' technical writing skills change in two classes with different TBL designs?
2. How does students' technical writing self-efficacy change in two classes with different TBL designs?
3. How do the female and male students' technical writing skills differ in two classes with different TBL designs?
4. How does the female and male students' technical writing self-efficacy differ in two classes with different TBL designs?

About the Case

This teaching comparison case was conducted in an undergraduate engineering course at a

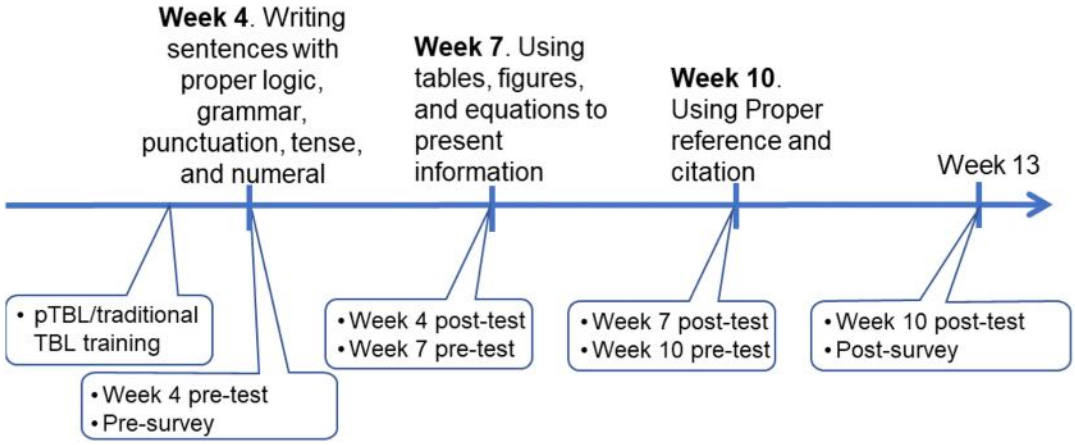


Fig. 1. Timeline of the technical writing sessions and pretests and posttests.

four-year university in the southern US. Students participated in this weekly 1-hour lab course during the semester. The course aimed to provide hands-on opportunities to understand the characteristics of different materials used in civil engineering. Students conducted lab experiments and wrote lab reports in small groups after each experiment. Technical writing sessions were offered before a lab experiment started.

Altogether, four 75-minute technical writing sessions were offered in the course (see Fig. 1). The first session, held during Week 4 of the semester, focused on writing sentences with proper logic, grammar, punctuation, tense, and numeral use. The second session, held during Week 7, covered the use of tables, figures, and equations to present information. The third session, held during Week 10, addressed proper reference and citation use. The fourth session, held during Week 13, focused on ethics and conducting a literature review. Only the first three sessions were included in this study, as the fourth did not have a post-test measurement because of its timing near the end of the term.

Two classes participated in this study, with each having a different intervention. The course instructor, one of the authors, taught both classes at the same pace with the same content.

Traditional TBL Class In the control group, students followed traditional TBL. Before Week 4, a practice session was held in class to familiarize students with the expectations and procedure of TBL. Starting in Week 4, students were assigned to teams of three. Team members were the same for the rest of the semester. The course did not have a technical writing textbook, but students were assigned readings before each session.

At the beginning of each session, students took an individual readiness assurance test, which consisted of six multiple-choice questions of knowledge retention. After that, the course instructor gave a 15-minute presentation. Following the lecture, students were asked to complete the readiness assurance test in small teams. Answers were discussed in class with the course instructor's facilitation. Occasionally, some teams appealed their answers. The course instructor offered explanations at this time.

In the application stage, teams were assigned authentic tasks to apply the knowledge that they had just acquired. For example, in Week 7, teams were given the raw data on the compressive strength values of concrete cylinders on different days. Those data were collected from one of the course instructor's previous projects. Students were required to draw a table, plot a figure, and write a paragraph explaining the data. The teams had about 25 minutes to finish and submit the application activities. Then their answers were exchanged across teams for peer evaluation. Teams wrote down three areas for improvement and three best performance areas before they nominated the best team. At the end of the TBL session, students were given a rubric to evaluate their team members' contributions to teamwork.

Peer TBL Class Students in the pTBL class followed the same TBL procedure as the traditional TBL class except for peer leadership. In Week 3, students received guidelines for how the leader and nonleader team members should behave in collaborative learning activities. A practice session was hosted that week for students to understand the expectations and procedure.

In Week 4, the course instructor assigned students into teams of three and randomly appointed a member as the team leader for each team. After the individual test and instructor's short lecture, students discussed and submitted a single test as a team. During the discussion, the student leaders sought each member's answers and explanations, kept team discussions on track, and facilitated collective decisions. They also appealed their answers in class if the team members agreed to do so.

In the application stage, pTBL students did the same activities as their traditional TBL peers except that student leaders facilitated the team's work. The leader asked questions that led to a discussion, probed in-depth comments or ideas, and gave members time to think and respond. The pTBL activities continued in Weeks 7 and 10, where the leaders were randomly selected from those who had not taken the leadership role in prior weeks. By the end of Week 10, all group members had been a peer leader at least once. Each pTBL session also ended with a peer evaluation, where each member used a rubric to assess others' contributions to teamwork.

Situating the Case

A variety of instructional strategies has been examined at the class level to improve students' technical writing skills. Bonham et al. [26] implemented a model among a group of physics students. The model contained the instructional elements of sample reports, fading scaffolding, and the instructor's ongoing feedback. Students reported improved writing skills and their favored attitudes regarding sample reports and instructor's feedback. Colton and Surasinghe [27] taught a technical writing class where students were assigned different roles and peer-reviewed each other's work in the small group writing. Results of the postcourse evaluation showed that students appreciated the interaction with their peers. Small-group learning was also adopted in Virtue's [28] class, in which students were assigned to work with group moderators. Students in the small-group class were found to have more interactions than those in the traditional instructor-led class. They felt comfortable and accountable in group discussions. Overall, the results of these studies indicate that students benefit from a technical writing class that includes elements of instructor's feedback, peer interaction, and opportunities for application.

TBL The TBL method orchestrates these elements into a well-structured collaborative learning approach. According to Michaelsen et al. [9], this approach begins with individual student preparation, such as reading assignments and watching videos before the class when the TBL activity will occur. Students' first activity in a TBL class is the individual readiness assurance test. Next, students work in teams to discuss their answers to that test. This activity is called the team readiness assurance test. The individual and team readiness assurance tests ensure that students have read the course materials before class and have the foundational knowledge needed to apply the course content.

After the team readiness assurance test, the instructor offers students the opportunity to appeal questions missed as a team. The instructor then clarifies confusions and misconceptions. The remainder of the module focuses on 4S application activities. Teams are given the Same Significant problems that are meaningful and require collective efforts to solve. Team members collaborate to select a Specific choice answer, which means that each problem is designed to have a single and clearly defined answer. All teams report their answers Simultaneously. The instructor facilitates the class discussion with teams defending their selection and articulating rationales for the answers. Although the readiness assurance process aims to ensure students' conceptual understanding, the 4S application activities require students to apply their knowledge to solve problems at higher cognitive levels.

TBL may wrap up with a peer evaluation activity, where students evaluate their group members' contributions.

TBL has been investigated in many studies, mainly in the medical and health-related areas [10], [11], [29]. Loftin and West [30] compared the outcome of physician-assistant students from TBL and lecture-based classes. They found that students in the TBL class outperformed their peers in the lecture-based class in terms of both academic performance and self-efficacy. Their study was confirmed by Swanson et al. [31], who conducted a meta-analysis and narrative synthesis of 30 studies in undergraduate and graduate education. They found that TBL had a medium and positive impact on students' content knowledge.

Despite the large number of TBL studies in medicine and health sciences, very few TBL studies have been conducted in undergraduate engineering education. When Parappilly et al. [32] implemented

TBL in a cohort of engineering students, they found that it narrowed the achievement gap between low- and high-performing students. However, Wang and Mott's [33] study rendered different results. They used the TBL approach in an introductory engineering course and found no significant difference in academic performance between students in the TBL and traditional instructor-led classes. In addition, they did not find that students favored TBL more than lectures. Students reported feeling lost in their self-study without instructors' presentations at the beginning of TBL.

In a nutshell, the literature has shown that although the impact of TBL in the medical and health-related areas was promising, studies in engineering education focused on students' academic performance were not conclusive. Besides, self-efficacy, a factor directly influenced by students' learning experience, was ignored in those studies. According to Bandura's [34], [35] self-efficacy theory, there is a relationship between students' belief in their capacity to complete a task and their performance. Studies have found that participants' self-efficacy was affected by their performance in the prior learning activities [36], [37]. At the same time, self-efficacy influenced participants' performance in upcoming learning activities. In short, our review of TBL literature indicates a need for additional studies to reveal its impact on engineering students' academic performance and self-efficacy.

Peer Leadership Chowdhury and Murzi summarized the optimal design of collaborative learning into 11 categories in their systematic reviews of engineering education [13], [38]. They included shared goal and value, motivation, interpersonal skills, open communication, constructive feedback, group composition, leadership, accountability, commitment to group success, and interdependence and adherence to group process and performance. Most of these categories were addressed in our TBL design except for leadership. Fink [39] explained that role assignment might prevent the organic development of group dynamics and students' accountability. Therefore, he did not recommend that it be used in TBL.

However, Chowdhury and Murzi [13] asserted that leadership was a crucial factor in effective teamwork. Peer leaders generally monitor group task completion, provide regular feedback to members, and ensure equal participation. Chowdhury and Murzi's [13] conclusion was confirmed in peer-led learning studies [21], [40]. For example, Everett et al. [24] assigned student leaders in small groups in engineering classes. They found that when more students had a chance to take the leading role, their learning outcomes improved.

Gendered Issue The student population in engineering classes is usually male-dominated [41]. Gendered behavior was found to affect students' collaborations in this discipline [16], [17]. In Wolfe and Alexander's [17] technical writing class, male students controlled the computers but avoided writing, revision, or peer-critique activities. As a result, computer tasks were highly valued in teams but writing tasks were not. However, in a qualitative study by Ingram and Parker [16], no consensus was identified regarding the gender difference in collaborative learning. They found that female students displayed different interaction styles depending on the roles and tasks that they took. Based on their literature review and observation, they suggested that collaborative learning in engineering classes should be designed to encourage women to assert themselves and to encourage men to collaborate.

Peer leadership seems to be a solution to the gendered issue [22], [42]–[45]. For example, Drane et al. [43] conducted a longitudinal study in an engineering course over 10 years. They examined the impact of a peer-led small-group learning program among undergraduate students. The program hired undergraduates as peer facilitators. They had done well in the same courses in prior semesters and were trained in group facilitation skills. Once or twice each week, peer facilitators led a group discussion on problem-solving and actively sought participation from every group member. The researchers found that the program benefited both female and male students. The course-grade effect was greater on men than women; however, that difference was not statistically significant. This finding was supported by other peer-led learning studies in chemistry, biology, and mathematics undergraduate education [22], [23], [46]. In those studies, peer leaders performed similar roles, such as promoting group members' participation and leading the group in problem-solving activities. No gender difference was detected in students' academic performance. Instead, peer-led learning was found equally effective for both genders. Based on our literature review, we posit that when peer leaders promote equal participation and collective decision-making, group members are encouraged to participate in activities previously dominated by

a single gender, and that encouragement improves their self-efficacy and learning outcome.

In summary, our literature review indicates that a quality technical writing class should use TBL that includes peer interactions, instructor feedback, and activities that target learning at different cognitive levels. This method could be enhanced with peer leadership, which was suggested in prior studies to narrow the gender difference in engineering students' learning.

METHODS/APPROACH

Participants Sixty-six undergraduate students who took the course were recruited for the study, 32 from the experimental group (the pTBL class) and 34 from the control group (the traditional TBL class). In the pTBL class, 81% identified themselves as males and 19% as females; 81% were Caucasians; and 3% were African American, American Indian, and Asian, respectively. Juniors comprised 72% of the respondents, and the rest were senior students. In the traditional TBL class, 65% self-identified as males; and 35% as females; 65% were Caucasians 15% were African Americans, 3% were Latino, and 6% were Asian. Juniors comprised 76%, and the rest were senior students.

Instruments and Data Collection Two sets of instruments were used in this teaching case study.

Technical Writing Skills: The course instructor developed three pairs of pretests and post-tests (see Fig. 1). The pretests were delivered at the beginning of each technical writing session in Weeks 4, 7, and 10. The post-tests were provided before the subsequent sessions were held in Weeks 7, 10, and 13. Each pair of pretests and post-tests used the same five multiple-choice or multiple-answer questions, and assessed students' technical writing skills covered in the week when the pretest was delivered. Each question was worth 5 points. Students had 10 minutes to complete each test.

The tests assessed students' lower and higher cognitive levels of knowledge. According to Bloom's taxonomy [47], lower cognitive level questions require students to recall and understand a term definition (e.g., Which of the following is incorrect about plagiarism? in Week 10). Higher cognitive level questions require students to analyze a problem, draw connections among ideas, and select an appropriate solution (e.g., If you conducted a slump test following the ASTM C143 standard and cited ASTM C143 in the text of your report, how should it appear in your reference list? in Week 10).

Self-Efficacy: We adopted and refined questions validated in Nardo and Hufana's [48] and Isnin's [49] studies to assess the self-efficacy of students' technical writing skills. We removed the items that were not taught in the TBL sessions. Altogether, 13 items were adopted in the study (see Table I). These items were included in the presurveys and postsurveys. The presurvey also included demographic questions and was delivered in Week 3. The postsurvey was delivered at the end of the study in Week 13. Students had 10 minutes to complete each survey in class.

Analysis of Overall Technical Writing Skills Because the data did not meet the normality assumption required for ANOVA tests, nonparametric tests were administered in this study. The IBM SPSS statistics software version 23 was used in the analyses. Three Wilcoxon tests were conducted in two classes, respectively. They were intended to identify significant changes in students' technical writing skills in the pretests and post-tests for Weeks 4, 7, and 10. In each test, the variables were the pretest and post-test scores of the relevant weeks.

Analysis of Overall Self-Efficacy Cronbach's alpha values were calculated for the 13 self-efficacy items in both the presurveys and postsurveys. Results showed excellent internal consistency, $\alpha_{\text{presurvey}} = 0.91$, $\alpha_{\text{postsurvey}} = 0.95$. Hence, means were calculated and used to represent students' technical writing self-efficacy before and after TBL sessions.

Then, a Wilcoxon test was administered for each class to examine the changes in students' self-efficacy before and after the technical writing sessions. In both tests, the variables were the means of self-efficacy scores that students reported in the presurveys and postsurveys.

Analysis of Gender Difference in Technical Writing Skills Mann-Whitney U tests were administered to compare the gender difference of students' technical writing skills in the two classes. The tests were operated separately in each class. In each test, the between-group variable was gender, and the within-group variables were students' post-test scores of the relevant weeks.

Analysis of Gender Difference in Self-Efficacy Two Mann-Whitney U tests were conducted to

TABLE I
ITEMS MEASURING TECHNICAL WRITING SELF-EFFICACY

Items		Source
1.	I am able to write a project report using the correct format.	[49]
2.	I am able to write technical documents with correct grammar.	[49]
3.	I am able to write technical documents with correct spelling.	[49]
4.	I am able to write technical documents with correct capitalization (capital and small letters).	[49]
5.	I am able to write technical documents with correct punctuation.	[49]
6.	I am able to illustrate a process.	[48]
7.	I am able to construct figures that present data clearly and precisely.	[49]
8.	I am able to construct tables that present data clearly and precisely.	[49]
9.	I am able to interpret graphic presentations such as figures and tables.	[49]
10.	I am able to write references for a project report using a correct way.	[49]
11.	I am able to differentiate the features of technical reports.	[48]
12.	I am able to spot errors in a technical report.	[48]
13.	I am able to distinguish the differences between technical writing and other forms of writing.	[49]

TABLE II
STUDENTS' TECHNICAL WRITING SKILLS IN THE PRETESTS AND POST-TESTS

		Week 4			Week 7			Week 10		
		N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Traditional TBL Class	Pre-test	31	15.87	3.39	32	13.34	4.85	31	16.68	4.36
	Post-test	31	17.58	4.26	32	15.62	6.06	31	16.61	5.83
pTBL Class	Pre-test	31	15.71	4.22	26	15.00	4.19	22	18.68	4.62
	Post-test	31	20.65	5.44	26	18.65	4.14	22	20.00	5.12

compare the self-efficacy between female and male students in the two classes. In both tests, the between-group variable was gender. The within-group variable was students' self-efficacy in the postsurvey.

RESULTS/DISCUSSION

This teaching case involved three technical writing sessions from two undergraduate civil engineering lab classes at a southern US university. Students participated in regular TBL activities in the technical writing sessions of one class, and their peers in the other class rotated the leadership role in the TBL activities. We compared students' technical writing skills and self-efficacy between the two classes. Tests of technical writing skills measured students' conceptual knowledge as well as their application of relevant technical writing concepts or principles acquired in each week. Students' self-efficacy was measured by their perceived ability to use technical writing skills. In addition, we examined the gender influence on their technical writing skills and self-efficacy in two classes.

Overall Technical Writing Skills Results of the Wilcoxon tests did not demonstrate a significant change of students' technical writing skills in the traditional TBL class, $Z_{\text{week4}} = -1.84$, $p_{\text{week4}} = 0.07$, $Z_{\text{week7}} = -1.76$, $p_{\text{week7}} = 0.08$, $Z_{\text{week10}} = -0.10$, $p_{\text{week10}} = 0.92$ (see Table II). However, students' technical writing skills in the pTBL class were significantly improved in Weeks 4 and 7, but not Week 10, $Z_{\text{week4}} = -3.58$, $p_{\text{week4}} < 0.01$, $Z_{\text{week7}} = -3.62$, $p_{\text{week7}} < 0.01$, $Z_{\text{week10}} = -1.02$, $p_{\text{week10}} = 0.31$ (see Table II).

Overall Self-Efficacy Results of the Wilcoxon tests showed a significant improvement of students' self-efficacy in the traditional TBL class, $Z = -3.83$, $p < 0.01$ (see Table III). A significant

TABLE III
STUDENTS' TECHNICAL WRITING SELF-EFFICACY IN THE PRESURVEYS AND POSTSURVEYS

		N	Mean	*SD*
Traditional TBL Class	Pre-survey	34	4.06	0.47
	Post-survey	31	4.44	0.50
pTBL Class	Pre-survey	32	4.10	0.47
	Post-survey	32	4.39	0.40

improvement was also seen in the pTBL class, $Z = -2.75$, $p < 0.01$.

Gender Difference in Technical Writing Skills No gender difference was detected in the pTBL class in all three post-tests, $U_{\text{week4}} = 44.00$, $p_{\text{week4}} = 0.16$, $U_{\text{week7}} = 44.50$, $p_{\text{week7}} = 0.72$, $U_{\text{week10}} = 21.50$, $p_{\text{week10}} = 0.13$ (see Table IV). Likewise, students in the traditional TBL class did not show a gender difference on their Week 4 post-test, $U = 65.50$, $p = 0.07$. However, female students were found to perform significantly better than their male classmates in Weeks 7 ($U = 46.00$, $p < 0.01$) and 10 ($U = 66.00$, $p < 0.05$) post-tests in the traditional TBL class.

Gender Difference on Self-Efficacy Results of Mann-Whitney U tests of students' self-efficacy in the postsurvey did not show a significant gender difference in either the traditional TBL ($U_{\text{Traditional}} = 64.00$, $p = 0.09$) or the pTBL ($U_{\text{pTBL}} = 64.00$, $p = 0.98$) class (see Table V).

Discussions Findings of our teaching case confirm those of prior studies that TBL infusing peer leadership improves students' self-efficacy and their skills in writing appropriate sentences and using tables, figures, and equations [27], [28]. It is noteworthy that the post-tests were administered three weeks after the relevant technical writing sessions were completed. Hence, they measure students' retained skills rather than the immediate outcome of TBL. The quantitative analyses did not demonstrate any significant improvement in students' technical writing skills in the traditional TBL class. However, students in the pTBL class showed significant improvement in the first two technical writing sessions, as predicted by the findings of prior studies [24], [40]. In addition, pTBL was as effective as traditional TBL in improving students' technical writing self-efficacy. Taking these findings into consideration, we believe that peer leadership can maximize and sustain the power of TBL.

A gender difference was not detected in the Week 4 technical writing session in the traditional TBL class. However, it did appear in the Week 7 and Week 10 sessions. Female students in the traditional TBL class performed better than their male classmates in the Week 7 and Week 10 sessions, echoing the results from Das et al.'s [50] TBL study, which found that female students outperformed their male peers on problem-solving skills. Das et al. [50] explained that the better performance could be due to female students' preference for team learning. However, the gender difference was not identified in the pTBL class in our case. The result in our pTBL class confirms and expands findings from other peer-led learning studies that peer leadership, when used in TBL, has the potential of bridging gendered performance [22], [43], [46].

More important, gender differences found in the two classes of our case affirm the necessity of further examining TBL in engineering undergraduate education. If TBL fit female students' learning styles better than those of their male counterparts, why was the gender difference not discerned in the pTBL class? In our case, gender difference was not identified in the first session in the traditional TBL class. Was it because of the topic of that class or the fact that team dynamics were not fully developed yet—or for other reasons? We look forward to further research revealing these answers.

CONCLUSION

Limitations One limitation of this study is the small sample size. It is probably the main reason why our data didn't meet the normality assumption and we had to use the nonparametric tests. It could also be the reason why we failed to find significant results in some tests. In addition, the study size aggravated the problem of the unequal distribution of female and male students, especially in the pTBL class. Therefore, we call for studies with a large sample size so that the data meet the assumptions of parametric tests and a representative number of female students can be obtained for future analyses. In an ideal situation of our case, a minimum of 30 female students in each class would be needed to expect a valid interpretation based on the normal distribution analysis [51], [52]. Given the male-to-female ratio in our case, we suggest that future studies should include a total of at least 245 students.

TABLE IV
FEMALE AND MALE STUDENTS' TECHNICAL WRITING SKILLS IN THE PRETESTS AND POST-TESTS

			Week 4			Week 7			Week 10	
			N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Traditional TBL Class	Male	Pre-test	20	15.25	3.40	20	11.65	4.30	19	14.95	3.15
		Post-test	20	16.50	4.32	20	14.00	6.20	19	15.53	4.38
	Female	Pre-test	11	17.00	3.23	11	16.73	4.29	11	19.36	4.95
		Post-test	11	19.55	3.50	11	19.55	2.70	11	20.00	5.00
pTBL Class	Male	Pre-test	24	15.79	4.55	20	15.00	3.89	17	18.24	4.82
		Post-test	24	19.79	5.61	20	18.50	4.01	17	19.12	5.37
	Female	Pre-test	6	15.5	3.39	5	14.00	5.48	4	20.25	4.57
		Post-test	6	23.33	4.08	5	18.00	4.47	4	23.75	2.50

TABLE V
FEMALE AND MALE STUDENTS' TECHNICAL WRITING SELF-EFFICACY IN THE PRESURVEYS AND POSTSURVEYS

			N	Mean	*SD*
Traditional TBL Class	Male	Pre-survey	22	4.10	0.41
		Post-survey	21	4.32	0.52
	Female	Pre-survey	12	3.98	0.58
		Post-survey	10	4.71	0.35
pTBL Class	Male	Pre-survey	26	4.08	0.50
		Post-survey	26	4.38	0.42
	Female	Pre-survey	6	4.21	0.27
		Post-survey	5	4.40	0.42

Gender distribution was not the same between the two classes. Because we randomly appointed classes as experimental and control groups, it was difficult to control the distribution of female and male students in each class. We suggest that future studies use classes with similar gender distribution to minimize the potential impact of unequal gender distribution on the learning outcome.

We conducted only quantitative analyses. Although these quantitative analyses help verify the relationships, they lack observational evidence on how and why those relationships occurred. Hence, we suggest that future studies including qualitative data, such as students' discourse in the TBL activities and analysis of students' lab reports, to validate or supplement the quantitative results. Other qualitative evidence such as the interviews and answers to open-ended questions in surveys would also provide a more holistic view of how the impact occurs.

We also did not cover the full range of technical writing skills. Due to time constraints, the course instructor picked the four categories of technical writing skills after weighing skill importance in the discipline, top mistakes observed in prior lab reports, and time. As a result, the skills taught in this teaching case were focused on a narrow spectrum of foundational skills. We suggest future studies applying TBL and pTBL to other rhetorical knowledge such as organization and genre, which are usually covered in a technical writing course. This approach would aid with generalizing the impact of TBL or pTBL impact in college technical writing classes.

Implications for Educators Based on the results of this teaching case and prior studies [25], [40], we propose that students have an equal opportunity to take the leadership role in TBL regardless of their previous course achievement or ranking. In our teaching case, pTBL improved students' self-efficacy and skills, such as writing well-structured sentences and representing information with tables, figures, and equations. Every team member was appointed the leader role at least once. It gave students, especially those who were usually not good at asserting themselves, an opportunity to speak out and have their voices heard. We believe that this random and equal assignment of leadership role boosted individual students' self-efficacy and enhanced their learning accountability, which enabled them to achieve significant learning gains in our case.

At the same time, we suggest that the random and equal assignment of the leadership role should be administered with two other actions. First, students need to be trained and gain practice in

the leadership responsibilities ahead of time. The training and practice in our case not only helped the instructor to discern and correct students' inappropriate behaviors but also clarified students' expectations. Second, the leadership responsibilities need to focus on team cohesion, accountability, and participation equity. Christianakis [53] pointed out that inequity was likely to occur in collaborative learning when students' responsibilities relied on their content expertise. High-performing students usually had more power and attention than their low-performing peers. In our teaching case, students were not expected to be content experts. They were told that no questions were silly questions in training. These two actions, namely, training with practice and leadership responsibilities focusing on equitable teamwork, gave pTBL students clear guidance to follow in our teaching case.

Implications for Researchers Our teaching case focused on using pTBL to assist students' learning of three sets of technical writing skills.

1. Writing sentences with proper logic, grammar, punctuation, tense, and numerals
2. Creating appropriate tables, figures, and equations to represent information
3. Citing and referencing others' work

The pTBL class observed students' improvement of the first two sets of skills. The learning of the first two sets of skills requires students to apply relevant rules and use their creativity to generate solutions, according to Welch [54]. In our opinion, equal participation in the pTBL class promotes the generation of creative ideas and discussion among team members. The collective intelligence refines the creative ideas and turns them into solutions. On the contrary, the third set of skills, citing and referencing others' work, consists of applying strict rules and guidelines with little room for creativity. Therefore, we posit that pTBL will benefit students more on topics allowing flexible and creative answers, such as the rhetorical aspect of technical writing [55]. We call for future studies examining the best fit of pTBL among different technical writing learning topics.

Studies of Bandura's self-efficacy theory have shown a mutual effect between students' self-efficacy and their technical writing skills [36], [37]. In our study, students in both the traditional TBL and pTBL classes significantly improved their technical writing self-efficacy, and we did not find any gender difference. In other words, both traditional TBL and pTBL improved female and male students' self-efficacy, a finding that may motivate them to take advanced technical writing classes in the future. Further investigation of when students build up their self-efficacy is worth pursuing. Qualitative research methods, such as interviews, are needed to supplement and further explain the quantitative results. Which technical writing session strengthens their self-efficacy? If we apply Bandura's self-efficacy theory, is students' self-efficacy affected by their most recent performance? How is the self-efficacy that was built up in earlier sessions sustained in the latter part of the course? Answers to these questions will help instructors understand the development of students' self-efficacy in the technical writing sessions and use appropriate strategies to promote their self-efficacy.

References

[1] B. Yalvac, H. D. Smith, J. B. Troy, and P. Hirsch, "Promoting advanced writing skills in an upper-level engineering class," *J. Eng. Educ.*, vol. 96, no. 2, pp. 117–128, Apr. 2007, doi: 10.1002/j.2168-9830.2007.tb00922.x.

[2] N. Gnanapragasam, "Evolution of technical writing in senior design—A case history," *Adv. Eng. Educ.*, vol. 2, no. 1, pp. 1–60, 2010.

[3] R. S. Weissbach and R. C. Pflueger, "Collaborating with writing centers on interdisciplinary peer tutor training to improve writing support for engineering students," *IEEE Trans. Prof. Commun.*, vol. 61, no. 2, pp. 206–220, Jun. 2018.

[4] D. E. Gragson and J. P. Hagen, "Developing technical writing skills in the physical chemistry laboratory: A progressive approach employing peer review," *J. Chem. Educ.*, vol. 87, no. 1, pp. 62–65, Jan. 2010, doi: 10.1021/ed800015t.

[5] D. Tatzl, W. Hassler, B. Messnarz, and H. Flühr, "The development of a project-based collaborative technical writing model founded on learner feedback in a tertiary aeronautical engineering program," *J. Tech. Writing Commun.*, vol. 42, no. 3, pp. 279–304, 2012, doi: 10.2190/TW.42.3.f.

[6] L. Melonçon, J. Rosselot-Merritt, and K. St. Amant, "A field-wide metasynthesis of pedagogical research in technical and professional communication," *J. Tech. Writing Commun.*, vol. 50, no. 1, pp. 91–118, 2019, doi: 10.1177/0047281619853258.

[7] J. Shen, S. R. Hiltz, and M. Bieber, "Learning strategies in online collaborative examinations," *IEEE Trans. Prof. Commun.*, vol. 51, no. 1, pp. 63–78, Mar. 2008.

[8] C. A. Bodnar and R. M. Clark, "Can game-based learning enhance engineering communication skills?" *IEEE Trans. Prof. Commun.*, vol. 60, no. 1, pp. 24–41, Mar. 2017.
[9] L. K. Michaelsen, A. B. Knight, and L. D. Fink, *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*. Sterling, VA, USA: Stylus Publ., 2004.
[10] R. J. Sisk, "Team-based learning: Systematic research review," *J. Nursing Educ.*, vol. 50, no. 12, pp. 665–669, Dec. 2011, doi: 10.3928/01484834-20111017-01.
[11] S.-N. C. Liu and A. A. Beaujean, "The effectiveness of team-based learning on academic outcomes: A meta-analysis," *Scholarship Teach. Learn. Psychol.*, vol. 3, no. 1, pp. 1–14, 2017, doi: 10.1037/stl0000075.
[12] L. K. Michaelsen and M. Sweet, "The essential elements of team-based learning," *New Directions Teach. Learn.*, vol. 2008, no. 116, pp. 7–27, Winter 2008.
[13] T. Chowdhury and H. Murzi, "Literature review: Exploring teamwork in engineering education," presented at the 8th Research in Engineering Education Symp., Making Connections, Cape Town, South Africa, Jul. 2019.
[14] K. Beddoes and G. Panther, "Gender and teamwork: An analysis of professors' perspectives and practices," *Eur. J. Eng. Educ.*, vol. 43, no. 3, pp. 330–343, 2017, doi: 10.1080/03043797.2017.1367759.
[15] J. Roy. (2019). Engineering by the numbers. [Online]. Available: https://ira.asee.org/wp-content/uploads/2019/07/2018-Engineering-by-Numbers-Engineering-Statistics-UPDATED-15-July-2019.pdf
[16] S. Ingram and A. Parker, "Gender and modes of collaboration in an engineering classroom: A profile of two women on student teams," *J. Bus. Tech. Commun.*, vol. 16, no. 1, pp. 33–68, Jan. 2002.
[17] J. Wolfe and K. P. Alexander, "The computer expert in mixed-gendered collaborative writing groups," *J. Bus. Tech. Commun.*, vol. 19, no. 2, pp. 135–170, Apr. 2005, doi: 10.1177/1050651904272978.
[18] D. M. Wilson, R. Bates, E. P. Scott, S. M. Painter, and J. Shaffer, "Differences in self-efficacy among women and minorities in STEM," *J. Women Minorities Sci. Eng.*, vol. 21, no. 1, pp. 27–45, 2015, doi: 10.1615/JWomenMinorScienEng.2014005111.
[19] G. S. Stump, J. C. Hilpert, J. Husman, W.-T. Chung, and W. Kim, "Collaborative learning in engineering students: Gender and achievement," *J. Eng. Educ.*, vol. 100, no. 3, pp. 475–497, Jan. 2013, doi: 10.1002/j.2168-9830.2011.tb00023.x.
[20] M. Laeser, B. M. Moskal, R. Knecht, and D. Lasich, "Engineering design: Examining the impact of gender and the team's gender composition," *J. Eng. Educ.*, vol. 92, no. 1, pp. 49–56, Jan. 2013, doi: 10.1002/j.2168-9830.2003.tb00737.x.
[21] J. R. Keup, "Peer leadership as an emerging high-impact practice: An exploratory study of the American experience," *J. Student Affairs Afr.*, vol. 4, no. 1, pp. 33–52, 2016, doi: 10.14426/jsaa.v4i1.143.
[22] R. F. Frey, A. Fink, M. J. Cahill, M. A. McDaniel, and E. D. Solomon, "Peer-led team learning in General Chemistry I: Interactions with identity, academic preparation, and a course-based intervention," *J. Chem. Educ.*, vol. 95, no. 12, pp. 2103–2113, Oct. 2018, doi: 10.1021/acs.jchemed.8b00375.
[23] A. Tenney and B. Houck, "Peer-led team learning in introductory biology and chemistry courses: A parallel approach," *J. Math. Sci., Collaborative Explorations*, vol. 6, no. 1, 2003, Art. no. 5, doi: 10.25891/6PGZ-J157.
[24] J. Everett, K. Mallouk, and J. Morgan, "Pre-defined roles and team performance for first year students," *Int. J. Learn., Teach. Educ. Res.*, vol. 16, no. 7, pp. 84–101, Jul. 2017.
[25] S. Zha and C. L. Ottendorfer, "Effects of peer-led online asynchronous discussion on undergraduate students' cognitive achievement," *Amer. J. Distance Educ.*, vol. 25, no. 4, pp. 238–253, Dec. 2011, doi: 10.1080/08923647.2011.618314.
[26] S. W. Bonham, K. Jones, B. Luna, and L. Pauley, "An integrated model for teaching writing in the introductory laboratory," *J. College Sci. Teaching*, vol. 48, no. 2, pp. 40–47, 2018, doi: 10.2505/4/jcst18_048_02_40.
[27] J. Colton and T. Surasinghe, "Using collaboration between English and Biology to teach scientific writing and communication," *J. College Sci. Teaching*, vol. 44, no. 2, pp. 31–39, 2014, doi: 10.2505/4/jcst14_044_02_31.
[28] D. Virtue, "Increasing student interaction in technical writing courses in online environments," *Bus. Prof. Commun. Quart.*, vol. 80, no. 2, pp. 217–235, 2017, doi: 10.1177/2329490617689880.
[29] P. Haidet, K. Kubitz, and W. T. McCormack, "Analysis of the team-based learning literature: TBL comes of age," *J. Excellence Coll. Teaching*, vol. 25, no. 3/4, pp. 303–333, 2014.
[30] C. Loftin and H. West, "Evaluating self-efficacy after a team-based learning activity," *J. Physician Assistant Educ.*, vol. 28, no. 2, pp. 96–102, Jun. 2017, doi: 10.1097/JPA.0000000000000119.
[31] E. Swanson, L. V. McCulley, D. J. Osman, N. S. Lewis, and M. Solis, "The effect of team-based learning on content knowledge: A meta-analysis," *Act. Learn. Higher Educ.*, vol. 20, no. 1, pp. 39–50, 2017, doi: 10.1177/1469787417731201.
[32] M. Parappilly, R. J. Woodman, and S. Randhawa, "Feasibility and effectiveness of different models of team-based learning approaches in STEMM-based disciplines," *Res. Sci. Educ.*, Oct. 2019, doi: 10.1007/s11165-019-09888-8.
[33] C. Wang and J. Mott, "Implementing team based learning in a first year introduction to engineering course," presented at the 122nd ASEE Annu. Conf. Expo., Seattle, WA, USA, 2015.
[34] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1986.
[35] A. Bandura, *Self-Efficacy: The Exercise of Control*. New York, NY, USA: Freeman, 1997.
[36] M. L. Bernacki, T. J. Nokes-Malach, and V. Aleven, "Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning," *Metacognition Learn.*, vol. 10, no. 1, pp. 99–117, Nov. 2014, doi: 10.1007/s11409-014-9127-x.
[37] K. Talsma, B. Schüz, R. Schwarzer, and K. Norris, "I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance," *Learn. Individual Differences*, vol. 61, pp. 136–150, 2018, doi: 10.1016/j.lindif.2017.11.015.

[38] H. G. Murzi, "Team-based learning theory applied to engineering education: A systematic review of literature," presented at the 121st ASEE Annu. Conf. Expo., Indianapolis, IN, USA, 2014.
[39] L. D. Fink, "Beyond small groups: Harnessing the extraordinary power of learning teams," in *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, L. K. Michaelsen, A. B. Knight, and L. D. Fink, Eds. Sterling, VA, USA: Stylus Publ., 2004, pp. 3–26.
[40] J. W. Bouw, V. Gupta, and A. L. Hincapie, "Assessment of students' satisfaction with a student-led team-based learning course," *J. Educ. Eval. Health Professions*, vol. 12, Jun. 2015, Art. no. 23, doi: 10.3352/jeehp.2015.12.23.
[41] K. White, S. K. Rumsey, and S. Amidon, "Are we 'there' yet? The treatment of gender and feminism in technical, business, and workplace writing studies," *J. Tech. Writing Commun.*, vol. 46, no. 1, pp. 27–58, 2015, doi: 10.1177/0047281615600637.
[42] M.-E. Cummings and I. Thompson, "Gender and empowerment in the technical writing class," presented at the 14th. Annu. Penn State Conf. Rhetoric Comp., University Park, PA, USA, 1995.
[43] D. Drane, M. Micari, and G. Light, "Students as teachers: Effectiveness of a peer-led STEM learning programme over 10 years," *Educ. Res. Eval.*, vol. 20, no. 3, pp. 210–230, Mar. 2014, doi: 10.1080/13803611.2014.895388.
[44] Y. Hartberg, A. B. Gunersel, N. J. Simspon, and V. Balester, "Development of student writing in biochemistry using calibrated peer review," *J. Scholarship Teaching Learn.*, vol. 2, no. 1, pp. 29–44, Feb. 2008.
[45] T. Forneris *et al.*, "Results of a rural school-based peer-led intervention for youth: Goals for health," *J. School Health*, vol. 80, no. 2, pp. 57–65, Feb. 2010, doi: 10.1111/j.1746-1561.2009.00466.x.
[46] X. Wu, "Peer-assisted learning in calculus II: Examining gender differences," presented at the 31st Semi-Annu. Honors and Undergraduate Research Scholars Poster Presentation, New York, NY, USA, 2019.
[47] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY, USA: Longman, 2001.
[48] M. T. B. Nardo and E. R. Hufana, "Development and evaluation of modules in technical writing," *Amer. J. Educ. Res.*, vol. 2, no. 6, pp. 341–350, 2014, doi: 10.12691/education-2-6-2.
[49] S. F. Isnin, "Exploring the needs of technical writing competency in English among polytechnic engineering students," *Int. J. Acad. Res. Bus. Social Sci.*, vol. 7, no. 12, pp. 77–90, 2017, doi: 10.6007/IJARBSS/v7-i12/3594.
[50] S. Das, K. Nandi, P. Baruah, S. K. Sarkar, B. Goswami, and B. C. Koner, "Is learning outcome after team based learning influenced by gender and academic standing?" *Biochem. Mol. Biol. Educ.*, vol. 47, no. 1, pp. 58–66, Jan. 2019, doi: 10.1002/bmb.21197.
[51] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-Step Approach*, 2nd ed. Hoboken, NJ, USA: Wiley, 2014.
[52] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, *Probability and Statistical Inference*, 9th ed. Boston, MA, USA: Pearson Educ., 2013.
[53] M. Christianakis, "'I don't need your help!' peer status, race, and gender during peer writing interactions," *J. Literacy Res.*, vol. 42, no. 4, pp. 418–458, 2010, doi: 10.1080/1086296X.2010.525202.
[54] K. D. Welch, "Poetry, visual design, and the how-to manual: Creativity in the teaching of technical writing," *English J.*, vol. 99, no. 4, pp. 37–42, Mar. 2010. [Online]. Available: http://www.jstor.org/stable/27807164
[55] D. Dysart-Gale, K. Pitula, and T. Radhakrishnan, "Improving professional writing for lay practitioners: A rhetorical approach," *IEEE Trans. Prof. Commun.*, vol. 53, no. 3, pp. 293–303, Sep. 2010.

Shenghua Zha received the Ph.D. degree in information science and learning technologies from the University of Missouri-Columbia, Columbia, MO, USA, in 2006. She is currently an Assistant Professor with the Department of Counseling and Instructional Sciences, University of South Alabama, Mobile, AL, USA. Her research interests include peer-led learning, small group collaboration, computational thinking, and learning strategies in a technology-enhanced environment. Her research has appeared in *Journal of Educational Computing Research, Journal of Digital Learning in Teacher Education*, *Telematics and Informatics*, and *Journal of Research on Technology in Education*.

Shenghua Wu received the Ph.D. degree in civil engineering from Washington State University, Pullman, WA, USA, in 2016. He is currently an Assistant Professor with the Department of Civil, Coastal, and Environmental Engineering, University of South Alabama, Mobile, AL, USA. His research interests include asphalt materials characterization, transportation sustainability and resiliency, pavement design, modeling and evaluation, recycling, rehabilitation and maintenance, as well as STEM education. He has authored/coauthored more than 40 peer-reviewed articles in journals such as JOURNAL OF CIVIL ENGINEERING EDUCATION, *Journal of Cleaner Production*, *Transportation Research Record*, *Construction and Building Materials*, and *International Journal of Pavement Engineering*.

Julie M. Estis received the B.S., M.S., and Ph.D. degrees in Communication Sciences and Disorders from the University of South Alabama, Mobile, AL, USA, in 1999, 2001, and 2004, respectively. She is the COVID-19 Response Coordinator and an Associate Professor with the University of South Alabama. With 16 years of teaching and research experience, her work extends from her previous role as a Director, Team USA Quality Enhancement Plan—a successful campus-wide initiative, which improved student learning, critical thinking, communication, and collaboration through team-based learning (TBL). As a TBL Collaborative Certified Trainer-Consultant, she is an experienced facilitator and consultant. She presents and publishes on the impact of TBL across various disciplines and instructional modalities to international audiences.

Book Review

Donald H. Cunningham, Edward A. Malone, and Joyce M. Rothschild

Technical Editing: An Introduction to Editing in the Workplace

—Reviewed by
ALAN HOUSER

Index Terms—*Copyediting, English grammar, proofreading, style.*

The authors of *Technical Editing* describe their primary audience as

> …students who are preparing themselves for technical communication careers and therefore need to understand and gain experience in the practice of technical editing. (p. x)

The authors also acknowledge secondary audiences of practicing technical communicators, professionals "in any field" who wish to improve their writing and editing skills, and writers and managers who wish to understand the contributions of technical editors and manage projects that include technical editing.

After an opening chapter on the history of technical editing, Chapter 2 "Preparing for an Editing Project" presents analysis of the rhetorical situation—the document's purpose, audience, and context. Chapter 3 "Planning and Implementing the Editing" addresses selecting the appropriate level of editing, and issues that focus on schedules and budgets.

Chapters 4–8 address editing for organization, navigation, completeness, accuracy, and style. Chapter 4 "Editing for Organization" provides common patterns of organization, and notably addresses that conflicting purposes may influence organization decisions.

Manuscript received October 18, 2021; accepted October 18, 2021. Date of publication November 11, 2021; date of current version November 24, 2021.
The reviewer is with Group Wellesley and Carnegie Mellon University, Pittsburgh, PA 15206 USA (email: arh@groupwellesley.com).

IEEE 10.1109/TPC.2021.3121527

Book publisher: New York, NY, USA: Oxford University Press, 2020, 578 pages, including index.

Chapter 5 "Editing for Navigation" addresses tactics for supporting users in navigating documents and finding information. The chapter addresses mechanisms for print navigation (tables of contents, lists of figures and tables, indexes), as well as navigation mechanisms for websites and other online formats. Online navigation aids discussed include navigation/menu bar design and placement, site maps, and navigational "breadcrumbs," which show the reader's context in the topic hierarchy of the website or other online content.

Chapter 6 "Editing for Completeness" and Chapter 7 "Editing for Accuracy" discuss (among other topics) how the audience's prior knowledge influences whether a document is "complete" and "accurate," as well as possible regulatory and safety considerations that may dictate document content.

Chapter 8 "Editing for Style" addresses editing for clarity and readability, presenting guidelines such as avoiding ambiguity, favoring active over passive voice, favoring positive over negative forms, minimizing nominalizations, and eliminating redundancies and clichés.

Chapter 9 "Editing Visuals" categorizes the types of graphs and other graphics used in text, their typical roles and use cases, and the way that graphics should accompany the text. Chapter 10 "Editing Page Design" presents design principles including alignment, repetition, and contrast, and also covers issues of typography.

Chapter 11 "Editing for Reuse" is particularly distinctive in its relevance to modern publishing workflows that are increasingly *topic*-oriented: in which writers create individual information topics, usually working with a semantic markup language,

and often in the context of a content management system. Authoring is agnostic of the way that topics will be collected and organized (the final "document"), and the final publishing format (print/PDF, desktop, smartphone). The authors provide informed overviews of content management systems, popular XML-based semantic markup languages (DITA, Lightweight DITA), and considerations for writing and editing content that will likely be published in multiple contexts, final formats, and assemblies of topics.

Chapters 12–17 specifically address copyediting for grammar and punctuation. Chapter 12 "Copyediting: Principles and Procedures" describes the copyeditors' task and covers editing in both hardcopy and online formats, including procedures for using Microsoft Word's online editing features.

The authors' coverage of grammar in Chapters 13–17 is similar in depth and scope to that of many English grammar texts and resources, and is clearly presented in the context of the copyeditors' task. All concepts are motivated with examples of both incorrect and correct usage.

Chapter 18 "Proofreading" addresses the scope of the proofreading task (particularly in contrast with the copyediting task) and addresses proofreading in both hardcopy and online contexts.

The book includes frequent sidebars that define key terms, provide commentary from practicing editors and writers, address historical context, explore ethical issues, or present possibly contrarian views. For example, the sidebar "Do We Really Need to Teach Students to Edit Hardcopy Anymore?" argues the value of the skill and practice of editing documents in hardcopy form (p. 296).

Throughout the book, the authors address issues of modern technical communication workflows. They acknowledge that the dedicated editor role has become increasingly rare and that many technical writers are responsible for editing their own or their peers' documents. They also address workplace considerations, including people, budgets, and schedules, in both planning and executing editing projects. Among their cautions:

> Writers and project managers can get upset if you move beyond proofreading and begin to rewrite passages that have already passed muster with the copyeditor. (p. 475)

The book includes a thorough 30-page glossary of grammar terms, 33 pages of notes and references, and a thorough index.

Subtitling this 578-page book as an "introduction" may be a misnomer; the book is broad in scope and deep in coverage of its subject. It fills a need for current texts on the topic of technical editing and is grounded in modern technical communication workflows, practices, and approaches. This book is an invaluable teaching aid for classrooms and a welcome reference resource for practicing professionals.

Book Review

George F. Hayhoe and Pam Estes Brewer

A Research Primer for Technical Communication: Methods, Exemplars, and Analyses, 2nd Edition

—Reviewed by
ALAN HOUSER

Index Terms—*Research methods, statistics, surveys, usability testing.*

The goals and purpose of *A Research Primer for Technical Communication* are appropriately summarized by the authors' words in the preface:

> …this book contains essential information that you need to know to perform—or be an informed consumer of—research in the field of technical communication.

The book is appropriate for advanced undergraduate and graduate courses that cover research methods, and for academics and practitioners with interest in reading, using (to inform decision-making), or contributing to technical communication research.

A Research Primer for Technical Communication covers the research methodologies appropriate to advance the field of technical communication, and provides adequate depth to appropriately evaluate or conduct technical communication research. Throughout the book, the authors present practical considerations for designing and executing technical communication research.

Although the understanding and application of statistical measures are critical for evaluating research results, the book is accessible to readers without a statistics background. The text does not generally present actual formulas for statistical measures. It does, however, provide plain-language descriptions of these statistical measures, their purpose in assessing research results, and the interpretation of their values. The book provides procedures for computing these statistics in Microsoft Excel and the open-source jamovi statistical software.

The authors do suggest consulting a statistics textbook if the reader requires more in-depth knowledge of statistical measures, or finds a need to apply other statistical tools to assess results.

The book is organized into two parts. Part I, Methods (Chapters 1–7), introduces the importance of technical communication research, describes the research process, and provides details for each of the five research methodologies discussed.

Chapter 1, Research: Why We Do It and Why We Read It, introduces research concepts, including categories of research, categories of research methods, and sources (sponsors or hosts) of research. The chapter also presents the major types of research publications, distinguishing between open publications, editor-regulated publications, and refereed journals, with the characteristics and degree of influence of each.

Chapter 2, The Research Phases and Getting Started, introduces the structure of a research paper: the Introduction, Methodology, Results, and Discussion (IMRAD) pattern, and explains each element of the pattern. The chapter addresses issues of research participation, including consent of subjects and Institutional Review Board (IRB) approval. Chapter 2 concludes by introducing Richard Halley's RAD concept, a recurring

Manuscript received October 18, 2021; accepted October 18, 2021. Date of publication November 11, 2021; date of current version November 24, 2021. *The editor-in-chief, who is one of the authors of this book, did not read or otherwise approve this review prior to its publication.*
The reviewer is with Group Wellesley and Carnegie Mellon University, Pittsburgh, PA 15206 USA (email: arh@groupwellesley.com).

IEEE 10.1109/TPC.2021.3121526

Book publisher: New York, NY, USA: Routledge, 2021, 316 pp., including index.

assessment mechanism throughout the book. RAD-conforming research is Replicable (the research is presented in sufficient detail that others can repeat the study), Aggregable (results are sufficiently supported that they can be combined with other results to build a body of data), and Data-supported (not based on impressions or subjective assessment). The authors use the RAD concept throughout the book, both in the discussion of each research method, and in assessing the validity of the exemplar research studies.

Chapters 3–7 each cover the five primary methods of technical communication research: the literature review, quantitative studies, qualitative studies, surveys, and usability studies. The authors present guidance and principles for designing studies using each method.

Each chapter in Part II, Exemplars and Analysis (Chapters 8–12), reprints a recent article in each of the five genres, and provides commentary on the authors' approaches, methodology, results, and conclusions. Each paper is of high quality, providing an example of effective research in each genre. Papers are sourced from *Technical Communication*, IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, *Technical Communication Quarterly*, and the *Journal of Usability Studies*.

Throughout the book, the authors provide motivating exercises. A single appendix covers citation styles. The book includes an index of adequate detail for the topics and methods covered in the text.

A Research Primer for Technical Communication was first published in 2008 (the original edition was authored by Michael A. Hughes and George F. Hayhoe). Chapters 1–6 of the 2nd edition include minor organizational changes and minor additions to bring the text up to date with current tools and practices. For example, the 2nd edition mentions Google Scholar as a research tool and provides a mini-tutorial on using the jamovi open-source statistical software. The 2nd edition also adds a RAD assessment to the explanation of each research methodology, and to the analysis of each exemplar.

All exemplars are new in the 2nd edition. The 2nd edition includes a new chapter on usability studies (Chapter 7, Conducting Usability Studies), and new usability study exemplar (Chapter 12, Analyzing a Report on the Results of a Usability Study).

A Research Primer for Technical Communication offers a highly readable, nearly conversational style; the authors address the reader directly in the second person. The authors motivate every concept with an example, descriptive metaphor, or analogy. The text is sufficiently readable, approachable, informative, and enlightening to inform readers and evaluators of technical communication research. It will also inform and inspire a new generation of technical communication researchers.

Information for Authors

The IEEE Transactions on Professional Communication is an applied research publication sponsored by the Professional Communication Society (PCS) of the Institute of Electrical and Electronics Engineers (IEEE). Readers include engineers, scientists, technical and professional writers, information designers, managers, and others working as scholars, educators, and practitioners. All share an interest in the effective communication of technical and business information.

This Transactions includes these types of content: research articles, integrative literature reviews, case studies, tutorials, teaching cases, and book reviews. All content is available electronically at ieeexplore.ieee.org.

Editorial Policy

Material submitted to this TRANSACTIONS should not have been published elsewhere nor be under consideration elsewhere. (Substantially revised and expanded versions of papers from conference proceedings are acceptable.) When using previously published materials, authors are responsible for making all appropriate acknowledgments and disclaimers and for obtaining whatever permissions are needed to reprint any copyrighted materials.

The articles in this journal are peer reviewed in accordance with the requirements set forth in the *IEEE PSPB Operations Manual* (sections 8.2.1.C & 8.2.2.A). Each published article was reviewed by a minimum of two independent reviewers using a single-blind peer review process, where the identities of the reviewers are not known to the authors, but the reviewers know the identities of the authors. Articles will be screened for plagiarism before acceptance.

Corresponding authors from low-income countries are eligible for waived or reduced open access APCs.

Every effort is made to ensure double-blind review of submissions. Authors should take special care to ensure that no information in the manuscript identifies them or the organizations that they are affiliated with. The editorial staff has pledged to protect the proprietary nature of material submitted for publication. All authors of material accepted for publication must transfer copyright to the IEEE before publication by signing the copyright transfer form. There is no charge for publication in the IEEE Transactions on Professional Communication, nor is any payment made to contributors for the right to publish their work.

Preparing Articles for Submission

Authors should send one electronic copy of their manuscript to ieee-tpc@ieee.org. In MS Word format (.doc or .docx). Do not send files in other formats.

Format of Title Pages Your manuscript file should contain a title page (listing all authors' names, addresses, telephone numbers, and email addresses). There should be no identifying information on any other page of the manuscript. The second page of your manuscript file should provide the title, an informative abstract, and a list of index terms. Pages of manuscript files should be numbered.

More detailed guidance is available at the PCS website: http://sites.ieee.org/pcs/transactions-of-professional-communication/

Abstract The abstract must be a concise yet comprehensive reflection of what is in your article. In particular: The abstract must be self-contained, without abbreviations, footnotes, or references. It should be a microcosm of the full article. The abstract must be between 150-250 words. Be sure that you adhere to these limits; otherwise, you will need to edit your abstract accordingly. The abstract must be written as one paragraph, and should not contain displayed mathematical equations or tabular material. The abstract should include three or four different keywords or phrases, as this will help readers to find it. It is important to avoid over-repetition of such phrases as this can result in a page being rejected by search engines. Ensure that your abstract reads well and is grammatically correct.

References References to other sources should be numbered sequentially by order of mention in the text, with the number placed in brackets and printed on line (not as a superscript) like [1]. Only one reference should be listed per reference number. Footnotes for references or for other material may not be used. The list of all references cited in the text should appear in numerical order-of-mention as endnotes. See the examples that follow.

Book:
[1] K. S. Campbell, *Thinking and Interacting Like a Leader: The TILL System of Interpersonal Communication*. Chicago, IL: Parlay, 2006.

Article in a Periodical:
[2] I. Thompson, "Women and feminism in technical communication," *Journal of Business and Technical Communication*, vol. 13, no. 2, pp. 154–178, 1999.

Chapter in a Book:
[3] S. Kleinmann, "The reciprocal relationship of workplace culture and review," in *Writing in the Workplace: New Research Perspectives*, R. Spilka, Ed. Carbondale, IL: Southern Illinois University Press, 1993, pp. 56–70.

Paper Published in Proceedings:
[4] K. St. Amant, "Virtual office communication protocols: A system for managing international virtual teams," in *Proceedings of IEEE International Professional Communication Conference*, 2005, pp. 703–717.

Website:
[5] Structural Engineering Society—International. [Online]. Available: http://www.seaint.org

Electronic Journal Article:
[6] M. B. Sarkar, B. Butler, and C. Steinfield. (1995). Intermediaries and cybermediaries: A continuing role for mediating players in the electronic marketplace. *Journal of Computer-Mediated Communication*. [Online]. Available: http://jcmc.indiana.edu/vol1/issue3/sarkar.html

English Language Editing Services English language editing services can help refine the language of your article and reduce the risk of rejection without review. IEEE authors are eligible for discounts at several language editing services; visit the IEEE Author Center to learn more. Please note these services are fee-based and do not guarantee acceptance.

After Acceptance For Publication

All material accepted for publication is edited to conform to IEEE Transactions on Professional Communication style. Efforts are made to allow authors to review their work after editing, but the final decision of suitability for publication is made by the Editor-in-Chief

All authors must register for an Open Researcher and Contributor ID (ORCID) prior to publication. See orcid.org for details.

Copyright Form and Biography Statements Upon acceptance for publication, authors will be required to submit one copy of the completed IEEE Copyright Form, with the first author's signature. This form can be obtained at www.ieee.org/web/publications/rights/copyrightmain.html. All authors must provide current contact information and brief (50–100 words) professional biography statements, including IEEE membership and rank (if applicable).

Artwork (Tables and Figures) All tables and figures should be mentioned in the text. Tables should be numbered consecutively with roman numerals [Table IV]. Figures should be numbered consecutively with Arabic numerals [Fig. 3]. Tables will be typeset by editorial staff; the table contents should be included in the electronic file containing the manuscript.

Figures are scanned and must be prepared in camera-ready form to fit in either a single column (3.5 inch/8.9 cm maximum width) or full page (7.25 inch/18.4 cm maximum width). Color figures are allowed for the PDF version of the material. However, **authors should ensure that their figures will be acceptable when printed in black and white**. Electronic files for figures should be submitted in TIFF, PS, or EPS file formats.

Shortly before publication, authors receive proofs for proofreading only; authors may not substantively edit their work at this stage, nor do they have page proof approval rights. Authors are mailed one copy of the issue in which their article appears. Reprints are available for a fee.

Open Access

This publication is a hybrid journal, allowing either Traditional manuscript submission or Open Access (author pays OA fee) manuscript submission. Upon submission, if you choose to have your manuscript be an Open Access article, you commit to pay the discounted $2,045 OA fee if your manuscript is accepted for publication in order to enable unrestricted public access. If you would like your manuscript to be a Traditional submission, your article will be available to qualified subscribers and purchasers via IEEE Xplore. No OA payment is required for Traditional submission.

Digital Object Identifier 10.1109/TPC.2021.3127618

IEEE PROFESSIONAL COMMUNICATION SOCIETY

EDITORIAL STAFF

ABSTRACT TRANSLATORS

EDITORIAL BOARD

Digital Object Identifier 10.1109/TPC.2021.3127620