Understanding User Preferences for Interaction Styles in Conversational Recommender System: The Predictive Role of System Qualities, User Experience, and Traits

Raj Mahmud
School of Computer Science
University of Technology Sydney
Sydney, NSW, Australia
raj.pgm@outlook.com

Shlomo Berkovsky
Macquarie University
Sydney, NSW, Australia
shlomo.berkovsky@gmail.com

Mukesh Prasad
Computer Science
University of Technology Sydney
Sydney, NSW, Australia
mukesh.prasad@uts.edu.au

A. Baki Kocaballi
School of Computer Science
University of Technology Sydney
Sydney, NSW, Australia
baki.kocaballi@uts.edu.au

Abstract

Conversational Recommender Systems (CRSs) deliver personalised recommendations through multi-turn natural language dialogue and increasingly support both task-oriented and exploratory interactions. Yet, the factors shaping user interaction preferences remain underexplored. In this within-subjects study ($N = 139$), participants experienced two scripted CRS dialogues, rated their experiences, and indicated the importance of eight system qualities. Logistic regression revealed that preference for the exploratory interaction was predicted by enjoyment, usefulness, novelty, and conversational quality. Unexpectedly, perceived effectiveness was also associated with exploratory preference. Clustering uncovered five latent user profiles with distinct dialogue style preferences. Moderation analyses indicated that age, gender, and control preference significantly influenced these choices. These findings integrate affective, cognitive, and trait-level predictors into CRS user modelling and inform autonomy-sensitive, value-adaptive dialogue design. The proposed predictive and adaptive framework applies broadly to conversational AI systems seeking to align dynamically with evolving user needs.

CCS Concepts

• **Human-centered computing** → **User models**; *Dialog systems*; *User studies*; **Interaction design theory, concepts and paradigms**; • **Information systems** → *Recommender systems.*

Keywords

Conversational Recommender System, Conversational AI, User Preference Modelling

This work is licensed under a Creative Commons Attribution 4.0 International License.
OZCHI '25, Sydney, Australia
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2016-1/25/11
https://doi.org/10.1145/3764687.3764722

ACM Reference Format:
Raj Mahmud, Shlomo Berkovsky, Mukesh Prasad, and A. Baki Kocaballi. 2025. Understanding User Preferences for Interaction Styles in Conversational Recommender System: The Predictive Role of System Qualities, User Experience, and Traits. In *37th Australian Conference on Human-Computer Interaction (HCI) (OZCHI '25), November 29–December 03, 2025, Sydney, Australia.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3764687.3764722

1 Introduction

Conversational Recommender Systems (CRSs) support users in discovering and selecting items through natural-language dialogue. Conceptually, CRSs comprise three core modules [34]. First, a dialogue interface module integrates natural language understanding (NLU) and natural language generation (NLG). The NLU component processes user utterances to extract intents, entities, and contextual cues, while NLG generates fluent, coherent responses aligned with conversation history. Second, a dialogue management (DM) module tracks the dialogue state and selects appropriate system actions, such as offering suggestions, requesting clarification, or adjusting interaction strategy. Third, a recommendation engine (RE) selects candidate items based on user preferences and context. While traditional CRSs rely on collaborative filtering or content-based retrieval, retrieval-augmented generation (RAG) systems enhance this step by integrating external knowledge bases and large language models (LLMs) [85].

Early CRSs often employed rule-based, *task-oriented* dialogue policies with system-led initiative to efficiently guide users toward relevant items (e.g., "Here are the best-rated laptops under $1000") [13, 63]. In contrast, *exploratory* interaction modes emphasise open-ended discovery, probing questions, and invitations to browse (e.g., "Would you like to explore some new tech trends or emerging product categories?"), drawing from research on interactive search and critiquing systems [53, 69]. More recently, LLM-powered CRSs have expanded this space by enabling dynamic elaboration, contextual questioning, and on-the-fly generation of diverse recommendations [17, 85]. These systems support both *task-oriented* and *exploratory* strategies, offering greater flexibility in tone, initiative management, and conversational depth [49, 90]. Consequently, recent studies have shifted focus from accuracy-centric optimisation to dialogue strategy design, adaptive user modelling, and UX-oriented evaluation frameworks [20, 32, 35, 83]. While these efforts highlight the importance of interaction quality in shaping user satisfaction, trust, and continued use, it remains unclear how users experience different interaction styles, particularly in relation to affective and cognitive dimensions, and how these preferences vary across individual traits. Although the contrast between *task-oriented* and *exploratory* dialogue strategies has inspired various system design approaches [12, 52], user preferences for these modes remain poorly understood. Clarifying when and why individuals favour each style is essential for tailoring interaction strategies to user expectations. While frameworks such as ResQue [65] and CRS-Que [35] effectively capture retrospective evaluations of satisfaction and trust, they do not anticipate a user's preferred interaction style in advance. Moreover, individual factors, including control preferences, prior experience with CRSs, and demographic traits, shape users' perceptions of the system and influence their interaction style preferences. However, their impact on strategy preference has not been examined systematically [22, 79]. To address these gaps, we investigate three research questions:

RQ1: Which user-perceived CRS qualities and UX dimensions predict interaction preference?

RQ2: What user profiles emerge based on CRS qualities, and how do they differ in interaction preference?

RQ3: Do user characteristics moderate interaction preferences in CRS conversations?

We make four contributions to the theory and design of CRSs. First, we introduce a controlled interaction-preference paradigm contrasting *task-oriented* and *exploratory* dialogue strategies, enabling systematic analysis of affective and cognitive factors. Second, we extend the CRS-Que framework [35] to predictive preference modelling by linking perceived system qualities and UX appraisals to interaction choices. Third, clustering analysis reveals five latent user profiles that can inform value-based, trait-sensitive personalisation strategies. Fourth, we identify moderation effects of age, gender, and control preferences, and challenge the assumption that instrumental qualities such as perceived usefulness and effectiveness uniformly favour task-oriented interactions. Although grounded in recommendation dialogues, the insights and adaptive heuristics developed here generalise to conversational AI systems that must dynamically align with users' informational and affective needs. The remainder of this article is structured as follows. Section 2 reviews related work. Section 3 outlines the study design. Section 4 reports the results. Section 5 discusses implications and limitations. Section 6 concludes.

2 Related Work

We organise the literature review into four strands: (1) interaction styles, (2) CRS qualities, (3) UX dimensions, and (4) user profiling. Each strand contributes to understanding how users perceive and prefer different styles of conversational interaction. The first subsection addresses high-level dialogue strategies.

2.1 Interaction Styles

While early work prioritised algorithmic improvements, recent studies have begun examining how interaction strategies affect user experience [34, 76]. Among these, the contrast between *task-oriented* and *exploratory* styles has gained particular attention. In this context, *task-oriented* interactions aim to fulfil user goals efficiently, using brief exchanges that minimise user effort and reduce cognitive load. These systems are particularly suited to users with well-defined intents, and are often evaluated based on completion rates, decision time, and transactional efficiency [89]. In contrast, *exploratory* interactions promote discovery by offering elaborative prompts, contextual suggestions, or reflective questions. They aim not only to deliver recommendations but also to surface latent needs, help users articulate vague or evolving goals, and refine preference structures through guided conversation. These strategies are rooted in principles from exploratory search [53, 82], preference elicitation [62, 64], and sensemaking theory [61, 70]. For example, a CRS assisting with laptop shopping might, in a *task-oriented* mode, list top-rated models after a single query specifying price or brand. In contrast, the same system operating in an *exploratory* mode might begin by asking whether the user intends to use the device for gaming, remote work, or academic purposes, and follow up with questions about screen preferences, past frustrations, or anticipated mobility needs. Similarly, in a travel domain, a task-oriented CRS might immediately show flight deals based on a destination query, whereas an exploratory system could suggest alternative destinations aligned with the user's mood, budget, or prior trips, then offer itinerary variations with local experiences. In education or upskilling contexts, an exploratory CRS might help users clarify goals by probing into their current career situation, prior coursework, or preferred learning style before suggesting a curated learning path. Through such layered exchanges, *exploratory* interactions scaffold user decision-making in ways that are reflective, adaptive, and context-sensitive. Although both styles have merits, they afford different experiential qualities. Task-oriented interactions typically yield speed, clarity, and transactional utility. Exploratory dialogues, while slower, support agency, reflection, and satisfaction in ill-structured decision tasks. Recent work has also introduced adaptive flows that adjust initiative based on user traits [19, 50, 69], but systematic modelling of how interaction preferences relate to user traits remains limited. Vakulenko et al. [80] proposed interactive storytelling to support exploratory search, while Lei et al. [48] introduced the Estimation–Action–Reflection framework to facilitate adaptive flows. Gao et al. [26] further emphasised the importance of balancing exploration and exploitation in CRS design. Despite these contributions, key questions remain about how perceived system qualities influence users' interaction preferences. The next section examines these qualities in detail.

2.2 CRS Qualities

Beyond recommendation accuracy, users evaluate how well a system explains its suggestions, adapts to feedback, maintains conversational flow, and introduces novel or unexpected options. These perceptions influence both expectations and satisfaction [38, 66]. The CRS-Que framework [35] builds on the earlier ResQue model

[65] by adapting its constructs specifically to conversational contexts, where dialogue structure and agent behaviour are critical. Key qualities identified include accuracy, novelty, explanation clarity, adaptability, and conversational quality [35]. Unlike momentary user experience reactions, these qualities represent user beliefs about what the system should deliver, rather than transient feelings during use. Prior research shows that users' preference of these qualities varies by domain and individual traits. For example, novelty and adaptability tend to be more valued in *exploratory* or hedonic contexts, while accuracy and efficiency are paramount in *task-oriented* contexts [32, 38]. Such distinctions suggest that CRS qualities are strong predictors of users' dialogue strategy choices, especially when users must choose between *task-oriented* and *exploratory* dialogue. Moreover, although Berkovsky et al. [4] studied user model mediation and personalisation techniques primarily in broader recommender system domains rather than CRSs specifically, their work provides foundational methods for integrating heterogeneous user data sources. Such integration supports dynamic adaptation of system qualities and behaviours to reflect individual differences. This capacity for adaptation is crucial for tailoring interactions that align with user preferences and contexts in CRSs. However, system qualities alone do not fully explain how users appraise their experiences. UX research shows that subjective evaluations, especially affective and cognitive appraisals, play a crucial role.

2.3 UX Dimensions

In this study, UX refers specifically to post-interaction appraisals that reflect users' cognitive and emotional evaluations of the experience. These appraisals are distinct from behavioural indicators such as time spent or action frequency. While behavioural data offer observable metrics of engagement, perceptual evaluations reveal underlying judgments shaped by users' expectations, system behaviour, and contextual goals [47, 54, 59]. UX perceptions influence outcomes such as trust formation, decision confidence, and continued system acceptance [5]. Affective dimensions such as enjoyment and surprise are particularly relevant in open-ended or elaborative conversations, where engagement and perceived intelligence of the system are crucial [17, 63]. In contrast, dimensions such as usefulness and perceived effectiveness are more salient in task-oriented interactions, reflecting the system's ability to provide practical support and relevant outcomes [40]. Recent findings by Yun et al. [87] show that affective tone and perceived helpfulness significantly influence interaction preferences in LLM-powered CRS dialogues, especially when users seek elaborative or emotionally intelligent responses. Similarly, Gajos et al. [25] demonstrate that adaptive interface behaviours tailored to user ability and context can enhance perceived control and user satisfaction, suggesting that adaptivity plays a critical role in shaping post-interaction UX. Although UX is typically measured after interaction, it is not merely a passive outcome. Rather, it constitutes a structured, reflective judgment that mediates preference formation and behavioural intention [23, 27]. These post-task evaluations serve as interpretive summaries that integrate momentary affect with broader task-related reasoning. Kocaballi et al. [39] provide a comprehensive taxonomy of UX factors in conversational systems, including emotion, hedonic quality, motivation, and frustration. Their work highlights the multidimensionality of UX and the importance of capturing both experiential and evaluative components. Consistent with this view, the present study treats UX perceptions as explanatory variables in preference modelling, testing whether users' retrospective evaluations help forecast their interaction preferences beyond what can be inferred from system quality priorities alone. These perceptual outcomes are also shaped by user-level traits. Thus, the next section turns to individual differences and profiling approaches in CRS design.

2.4 User Profiling

User preferences in CRSs are shaped by a range of individual differences, including demographic traits, cognitive styles, behavioural tendencies, and control expectations. A key concept in this context is value orientations, defined as stable patterns in users' priorities and expectations regarding system attributes and interaction styles. These orientations shape how users evaluate and respond to different CRS behaviours. Research indicates that grouping users by their attitudes, behaviours, and value orientations helps tailor system strategies more effectively to meet individual needs [44, 67, 79]. For example, some users prioritise novelty and detailed explanations, shaping their expectations about dialogue style and system initiative [74]. Beyond segmentation, individual characteristics also moderate how users respond to system features. Users with high digital agency or strong trust in technology often prefer adaptive, *exploratory* interactions, whereas those who favour clear, directive guidance tend to prefer *task-oriented* interaction. Aligning interaction patterns with these preferences is critical for user satisfaction and system acceptance [77]. Additionally, demographic factors such as age, language proficiency, and prior experience influence perceptions of agent tone, fluency, and usefulness [22, 30]. Notably, Berkovsky et al. [2–4] demonstrate the benefits of integrating diverse user data sources and employing persuasive personalisation techniques to enhance recommendation relevance and user satisfaction. Their work supports the development of adaptive user profiles that enable personalised and context-aware CRS interactions.

These four strands of research underscore the complexity of CRS evaluation and personalisation. While prior work has explored dialogue strategies, system attributes, UX outcomes, and user traits, few studies have integrated these components to predict how individuals choose between dialogue styles. Moreover, existing frameworks tend to focus on retrospective evaluation rather than preference forecasting. By combining CRS qualities, retrospective UX ratings, and trait-level profiling in a predictive model, our study advances a more adaptive and explanatory perspective on dialogue preference in LLM-powered CRS environments.

3 Methods

We employed a within-subjects experimental design in which participants sequentially experienced two scripted CRS dialogues: one *task-oriented* and one *exploratory* as described in Section 2.1. The dialogue excerpts are given in Appendix A. The dialogues were situated in an online apparel-shopping context and were carefully matched for informational content and length to isolate the stylistic contrast. Participants rated their experiences, prioritised eight CRS qualities, and provided individual-difference data, allowing us to

Table 1: Characteristic features differentiating *task-oriented* and *exploratory* interaction styles in scripted CRS dialogues.

Aspect	Task-oriented	Exploratory
Dialogue style	Concise, direct, goal-oriented	Elaborative, open-ended, suggestion-rich
System prompts	Minimal elaboration, rapid delivery	Reflective, contextual, inviting elaboration
User engagement	Decision-focused, minimal probing	Exploration-focused, encourages deliberation
Tone	Functional, task-driven	Conversational, empathic, exploratory
System behaviours	Highlights deals, delivers options quickly	Offers context, rationale and alternatives

model preferences based on system quality beliefs, UX appraisals, and demographic and usage traits.

3.1 Research Design

The defining characteristics of the *task-oriented* and *exploratory* interaction styles are summarised in Table 1. These styles were operationalised based on exploratory search theory [53], conversational dialogue design principles [69], and the CRS-Que framework [35], which together provide theoretical grounding for contrasting interaction modes in CRS dialogues. The use of scripted dialogues was a deliberate methodological choice to maintain experimental control and stimulus equivalence, enabling systematic examination of dialogue style effects on user preference. Scripted or synthetic interactions are widely employed in conversational AI and human–computer interaction research to facilitate within-subject comparisons while controlling for content and length, thereby reducing variability inherent in live interactions [5, 8, 69]. Although scripted dialogues do not capture dynamic turn-taking or adaptive responses, their controlled nature enhances internal validity and supports precise measurement of user evaluations [11, 55]. After exposure to both dialogue conditions, participants indicated their preferred interaction style and completed survey measures assessing prioritised system qualities, retrospective user experience evaluations, and individual characteristics. To validate these scripts, we conducted pilot testing with a small sample and solicited expert feedback. Reviewers confirmed that the *task-oriented* dialogue conveyed a goal-focused concise tone, while the *exploratory* version was perceived as elaborative and suggestion-driven.

3.2 Participant Recruitment and Data Collection

Participants ($N = 191$) were recruited from three online platforms such as Prolific ($n = 149$), Survey Swap ($n = 33$), and Meta ($n = 9$) to ensure a demographically diverse sample. Eligibility criteria required participants to be at least 18 years old and proficient in English. For Prolific recruits, this was verified through platform-level screening (minimum 95% approval rate and at least 100 prior completions); for other sources, proficiency was self-reported. No restrictions were placed on gender, education level, or geographic location. Participants completed a structured Qualtrics survey lasting approximately nine minutes and received $2.40 as compensation. To ensure data quality, we applied a multi-criteria screening protocol that excluded responses with unusually fast completion times, straightlining patterns, or excessive missing or inconsistent data [46, 56]. Following screening, $N = 168$ valid responses remained. Of these, $N = 139$ participants expressed a definitive interaction preference and were retained for inferential analysis.

Sample Size and Power Analysis. We conducted an *a priori* power analysis using G*Power 3.1 to estimate the minimum sample size required for our primary hypothesis tests using logistic regression. Assuming a medium effect size $f = 0.25$, a significance threshold of $\alpha = 0.05$, and statistical power of 0.80, the required sample size was estimated to be 128 participants. Our final analytic sample of $N = 139$ meets this threshold, providing sufficient sensitivity to detect medium-sized effects in both the logistic regression and moderation analyses [24, 31].

3.3 Survey Instruments and Measures

The survey consisted of six modules: participant consent, demographics, dialogue stimuli exposure, UX appraisals, CRS quality prioritisation, and usage-related variables. Demographic data included age group, gender, education level, occupation, and self-reported English language proficiency. As described in Section 3.1, participants experienced two matched scripted CRS dialogues differing only in interaction style (*task-oriented* vs. *exploratory*), allowing us to isolate stylistic effects. Interaction preference was assessed using a categorical forced-choice item, where participants indicated which of the two dialogues they found more desirable or preferable. The response options included preference for Dialogue 1 (task-oriented), Dialogue 2 (exploratory), both equally, none, or unsure. For analytical clarity, only participants indicating a clear preference for either Dialogue 1 or 2 were retained. These responses were recoded into a binary variable representing preference for the *task-oriented* (coded 0) or *exploratory* (coded 1) dialogue style. Neutral or ambivalent responses (both, none, unsure) were excluded from inferential analyses.

User experience was evaluated across four dimensions: enjoyment, surprise, usefulness, and perceived effectiveness. These dimensions were adapted from established human–computer interaction frameworks [28, 38] and were rated separately for each dialogue. Although measured on Likert scales, UX scores were treated as continuous variables and standardised prior to analysis, consistent with accepted psychometric practice [9, 58].

Participants also rated the importance of eight CRS qualities, derived from the CRS-Que framework [35], including accuracy, novelty, explanation clarity, adaptability, conversation quality, attentiveness, understanding, and response quality. These ratings served as independent predictors in regression analyses and as features for clustering models. All CRS quality ratings were standardised using z-scores. Participants reported their CRS usage frequency on an ordinal scale ranging from 'Never' to 'Always'. Control preferences regarding system-initiated versus user-initiated interactions were measured using five-point Likert-type items. For regression and moderation analyses, these variables were treated as continuous predictors following standard practice [58].

3.4 Data Analysis

All analyses were conducted in IBM SPSS and Python using pandas, statsmodels, and scikit-learn. Predictor variables measured on five-point Likert scales were standardised to z-scores prior to analysis to place them on a common metric and facilitate interpretation.

RQ1: Binary Logistic Regression. To model the probability that participant i preferred the exploratory interaction ($Y_i = 1$) over the task-oriented interaction ($Y_i = 0$), we applied binary logistic regression. This method is suitable for dichotomous dependent variables and enables interpretation of predictor effects in terms of log-odds, which is commonly used in behavioural research [31].

$$\operatorname{logit}\big(P(Y_i = 1)\big) = \beta_0 + \sum_{j=1}^{k} \beta_j \, x_{ij} \tag{1}$$

where $\operatorname{logit}(P)$ is the natural logarithm of the odds of preferring the exploratory interaction, β_0 is the intercept, β_j are regression coefficients, and x_{ij} are standardised predictor variables (e.g., CRS qualities and UX dimensions).

RQ2: Unsupervised Clustering. To identify latent user profiles, we applied k-means clustering to participants' standardised CRS quality ratings. This technique partitions users into K clusters by minimising within-cluster variance:

$$\min_{\{\mu_k\}} \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \|x_i - \mu_k\|^2 \tag{2}$$

where $r_{ik} = 1$ if participant i belongs to cluster k, and 0 otherwise. μ_k are cluster centroids. The solution was validated via hierarchical agglomerative clustering using Ward's method, with K selected via elbow/silhouette criteria and stability checks using adjusted Rand Index.

RQ3: Moderated Logistic Regression. To examine whether user traits (e.g., age, control preference) moderated the influence of CRS predictors on interaction preference, we extended the logistic model with interaction terms:

$$\begin{aligned} \operatorname{logit}\big(P(Y_i = 1)\big) = \beta_0 &+ \sum_{j=1}^{k} \beta_j x_{ij} + \sum_{m=1}^{M} \gamma_m z_{im} \\ &+ \sum_{j=1}^{k} \sum_{m=1}^{M} \delta_{jm} (x_{ij} \cdot z_{im}) \end{aligned} \tag{3}$$

where x_{ij} are CRS predictors, z_{im} are trait moderators, β_j, γ_m are main effects, and δ_{jm} are interaction coefficients.

Hypothesis Testing Criteria. Hypotheses were evaluated using a two-tailed significance threshold of $p \leq 0.05$, consistent with standard practice in behavioural research [14, 21]. A result yielding $p \leq 0.05$ was taken as sufficient evidence to reject the null hypothesis H_0 in favour of the corresponding alternative hypothesis H_1. Conversely, a result with $p > 0.05$ was interpreted as indicating insufficient evidence to reject H_0, without implying proof of its truth. This approach ensures that Type I error is controlled at $\alpha = 0.05$ while acknowledging that failure to reject H_0 reflects only a lack of decisive evidence against it given the sample data [81].

3.5 Research Hypotheses

RQ1: Predictive Role of CRS Qualities and UX Dimensions. The following hypotheses were defined to assess how system-level beliefs and retrospective experience appraisals predict interaction preference. Hypothesis H1a states that users' preference for specific CRS qualities (for example, novelty and adaptability) predicts their likelihood of preferring the *exploratory* interaction. Hypothesis H1b asserts that users' preference for qualities related to accuracy and response quality predicts their likelihood of preferring the *task-oriented* interaction. Hypothesis H1c proposes that higher retrospective ratings of enjoyment and surprise are positively associated with preference for the *exploratory* interaction. Finally, Hypothesis H1d posits that higher retrospective ratings of usefulness and perceived effectiveness are positively associated with preference for the *task-oriented* interaction.

RQ2: Latent User Profiles Based on Value Orientations. To investigate whether distinct user segments emerge from differential value orientations, Hypothesis H2a proposes that participants can be clustered into latent profiles according to their prioritised CRS qualities. Hypothesis H2b further predicts that these latent profiles will exhibit systematic differences in interaction preference.

RQ3: Moderation by Usage Patterns and Demographic Traits. The final set of hypotheses examines whether individual characteristics moderate the influence of CRS qualities on interaction preference. Hypothesis H3a predicts that CRS usage frequency will moderate the relationship between CRS qualities and preferred interaction style. Hypothesis H3b asserts that control preference (system-initiated versus user-initiated) will serve as a moderator of the same relationship. Hypothesis H3c proposes that demographic factors, specifically age and gender, will moderate users' interaction preferences in response to varying CRS qualities.

Study Planning and Hypotheses Transparency. The hypotheses and analytical plan were formulated *prior to data collection* and were guided by theoretical frameworks and relevant literature. Although the study was not pre-registered on a public platform, all hypotheses were defined a priori and directly aligned with our research questions. The explicit hypothesis statements in this section and their statistical outcomes in Table 7 reflect our commitment to analytical rigour and research transparency.

3.6 Ethics

The study protocol was reviewed and approved by the Human Research Ethics Committee of the University of Technology Sydney. All participants provided informed consent prior to participation. The study adhered to institutional guidelines for ethical research involving human subjects and complied with relevant data protection regulations, including the General Data Protection Regulation (GDPR). Participants' privacy and confidentiality were rigorously maintained. No personally identifiable information, such as names, geographic locations, or sensitive demographic details, was collected. Data were anonymised and stored securely in accordance with the University's data management policies.

Table 2: Logistic regression predicting preference for *exploratory* interaction from CRS qualities. OR > 1 = increased odds. Bold p-values indicate significance ($p \leq 0.05$). McFadden's $R^2 = 0.12$; $N = 139$.

Predictor	OR	95% CI	p
H1a: Positive association with exploratory interaction			
Novelty	1.66	[1.04, 2.63]	**.033**
Adaptability	0.79	[0.48, 1.31]	.366
Attentiveness	1.00	[0.59, 1.67]	.988
H1b: Positive association with task-oriented interaction			
Accuracy	1.11	[0.69, 1.77]	.674
Response quality	0.79	[0.47, 1.34]	.383
Other CRS qualities			
Explainability	1.57	[0.98, 2.52]	.062
Conversation quality	1.77	[1.04, 2.99]	**.035**
Understanding	0.70	[0.42, 1.16]	.163

Table 3: Logistic regression predicting preference for *exploratory* interaction from UX dimensions. OR > 1 = increased odds. Bold p-values indicate significance ($p \leq 0.05$). McFadden's $R^2 = 0.32$; $N = 139$.

Predictor	OR	95% CI	p
H1c: Affective UX dimensions			
Enjoyment	2.19	[1.37, 3.51]	**.001**
Surprise	0.68	[0.42, 1.10]	.119
H1d: Task-oriented UX dimensions			
Usefulness	2.23	[1.32, 3.77]	**.003**
Perceived effectiveness	2.17	[1.28, 3.69]	**.004**

4 Results

We organised this section sequentially to align with each research question and its associated hypotheses, beginning with predictive models of interaction preference (RQ1), followed by cluster-based user profiling (RQ2), and concluding with moderation analyses exploring the influence of individual characteristics (RQ3). A consolidated summary presenting the key results is provided in Table 7.

4.1 Predictors of Interaction Preference (RQ1)

Two binary logistic regression models were fitted: one using eight CRS quality attributes derived from the CRS-Que framework [35], and another using four UX dimensions—enjoyment, surprise, usefulness, and perceived effectiveness, as expressed in Equation 1. Both models were statistically significant, with McFadden's pseudo R^2 values of 0.12 and 0.32, respectively, indicating modest to moderate explanatory power. As shown in Table 2, higher preference for *novelty* (Odds Ratio [OR] = 1.66, 95% Confidence Interval [CI] [1.04, 2.63], $p = 0.033$) and *conversation quality* (OR = 1.77, 95% CI [1.04, 2.99], $p = 0.035$) significantly increased the odds of preferring the *exploratory* interaction. Preference for *accuracy* (OR = 1.11, $p = 0.674$) and *response quality* (OR = 0.79, $p = 0.383$) did not significantly predict preference for the *task-oriented* interaction. Explainability showed a trend toward significance ($p = 0.062$), while attentiveness, understanding, and adaptability were not significant predictors.

Table 3 presents results from the UX dimension model. Higher ratings of *enjoyment* (OR = 2.19, 95% CI [1.37, 3.51], $p = 0.001$), *usefulness* (OR = 2.23, 95% CI [1.32, 3.77], $p = 0.003$), and *perceived effectiveness* (OR = 2.17, 95% CI [1.28, 3.69], $p = 0.004$) were significantly associated with increased odds of preferring the *exploratory* interaction. The *surprise* dimension was not a significant predictor (OR = 0.68, 95% CI [0.42, 1.10], $p = 0.119$). These findings support Hypotheses H1a and H1c, indicating that users who prefer novelty, conversation quality, and positive affective UX components such as enjoyment are more likely to prefer the *exploratory* interaction. Hypotheses H1b and H1d were not supported, as preference for accuracy and response quality did not predict preference for the *task-oriented* interaction, and usefulness and perceived effectiveness unexpectedly related positively to *exploratory* preference. In summary, participants' preference for the *exploratory* interaction was more strongly associated with *novelty* and *conversation quality* than with *accuracy* or *response quality*. These results indicate that, when controlling for interaction style, users place greater emphasis on hedonic and conversational qualities rather than solely on instrumental attributes.

4.2 Cluster Profiles and Interaction Preference (RQ2)

We applied both k-means and hierarchical (Ward's method) clustering to participants' z-standardised ratings of eight CRS-Que qualities. Although both methods yielded similar high-level segment structures, we selected the hierarchical solution for reporting because it offered superior stability and more balanced cluster sizes as shown in Appendix Figure A.1 and Table A.1 [1]. Solutions for $k = 3$–9 were evaluated using combined elbow-method and silhouette-score diagnostics; the five-cluster solution ($k = 5$) achieved the best trade-off among within-cluster cohesion, between-cluster separation, stability (mean ARI = 0.770, SD = 0.214), and minimum cluster size ($n \geq 18$). All subsequent analyses focus on this five-cluster segmentation. The association between cluster membership and interaction preference was robust for the hierarchical solution ($\chi^2(4, N = 139) = 14.93$, $p = 0.0048$; Kruskal–Wallis $H = 14.82$, $p = 0.0051$). The average silhouette score (0.146) further confirmed the validity of the segmentation.

Cluster Characterisation. Table 4 reports the mean importance ratings for each CRS quality by cluster. Each column represents one user segment; the header row indicates cluster size. Cluster 1 ($n = 34$) consistently rated all qualities highly (means 4.41–4.91), while Cluster 2 ($n = 18$) showed moderate to low engagement (means 2.89–3.33). Cluster 3 ($n = 24$) prioritised adaptivity and response quality but rated novelty lower. Cluster 4 ($n = 26$) valued accuracy and adaptivity with moderate novelty and conversation quality. Cluster 5 ($n = 37$) emphasised novelty, adaptivity, and response quality, with balanced ratings elsewhere.

Interaction Preference Patterns. Table 5 shows how participants' interaction preferences (0 = task-oriented, 1 = exploratory) distribute across clusters. Cluster 1 exhibited a strong exploratory preference (23 of 34 participants). Cluster 2 was balanced (9 of 18).

Table 4: Mean importance ratings for CRS qualities across five user clusters (Ward's method, $k = 5$, $N = 139$). Sample sizes: C1=34, C2=18, C3=24, C4=26, C5=37. Scores range 1 (low) to 5 (high).

CRS Quality	C1	C2	C3	C4	C5
Accuracy	4.85	3.22	4.17	4.81	3.95
Explainability	4.62	3.11	3.71	3.65	3.51
Novelty	4.41	3.11	2.71	3.54	4.03
Conv. Qual.	4.79	3.22	2.96	3.81	3.92
Attentiveness	4.91	3.28	3.50	4.27	3.97
Understand.	4.74	3.17	3.96	4.77	4.03
Adaptability	4.68	2.89	4.29	4.19	4.32
Resp. Qual.	4.91	3.33	3.96	4.46	4.22

Table 5: Distribution of interaction preferences by user cluster ($k = 5$, $N = 139$). Each row shows the number of participants in a cluster who preferred either task-oriented (0) or exploratory (1) CRS dialogue.

Cluster	Task-oriented	Exploratory
C1	11	23
C2	9	9
C3	17	7
C4	18	8
C5	26	11

Clusters 3, 4, and 5 each favoured the task-oriented mode: Cluster 3 (17 of 24), Cluster 4 (18 of 26), and Cluster 5 (26 of 37). For a graphical overview of cluster value profiles, see Appendix Figure A.2.

4.3 Moderating Effects of User Characteristics (RQ3)

A series of moderated logistic regression analyses were conducted to examine whether the effects of perceived CRS qualities on interaction preference were systematically moderated by individual user characteristics, including usage frequency, control preference, age, and gender. The outcome variable was binary, reflecting preference for the *exploratory* interaction style. Eight CRS qualities were included as predictors, each standardised (z-scored). For each quality, interaction terms were specified with all four moderators. Significant interaction effects ($p \leq 0.05$) are summarised in Table 6. Six interaction terms were statistically significant at the $p \leq 0.05$ level: Explainability × Gender, Novelty × Control Preference, Conversation Quality × Age, Adaptability × Age, Response Quality × Usage Frequency, and Response Quality × Age. For all significant interactions, model-predicted probabilities of preferring the *exploratory* interaction style are visualised in Figure 1. In each panel, the probability of *exploratory* interaction preference is plotted as a function of the corresponding CRS quality (z-score), stratified by levels of the moderating variable. No other quality × moderator interactions reached statistical significance at $p \leq 0.05$.

Table 6: Significant interaction effects ($p \leq 0.05$) from moderated logistic regression predicting preference for the *exploratory* interaction mode. Only interactions with $p \leq 0.05$ shown.

CRS Quality	Moderator	Estimate	Std. Error	z	p
Explainability	Gender	−4.94	2.11	−2.34	**.019**
Novelty	Control Pref.	2.79	1.30	2.14	**.033**
Conv. Qual.	Age	−4.00	1.98	−2.02	**.044**
Adaptability	Age	−2.95	1.50	−1.96	**.050**
Resp. Qual.	Usage Freq.	1.53	0.75	2.04	**.041**
Resp. Qual.	Age	3.28	1.37	2.40	**.016**

4.4 Summary of Hypotheses Tests

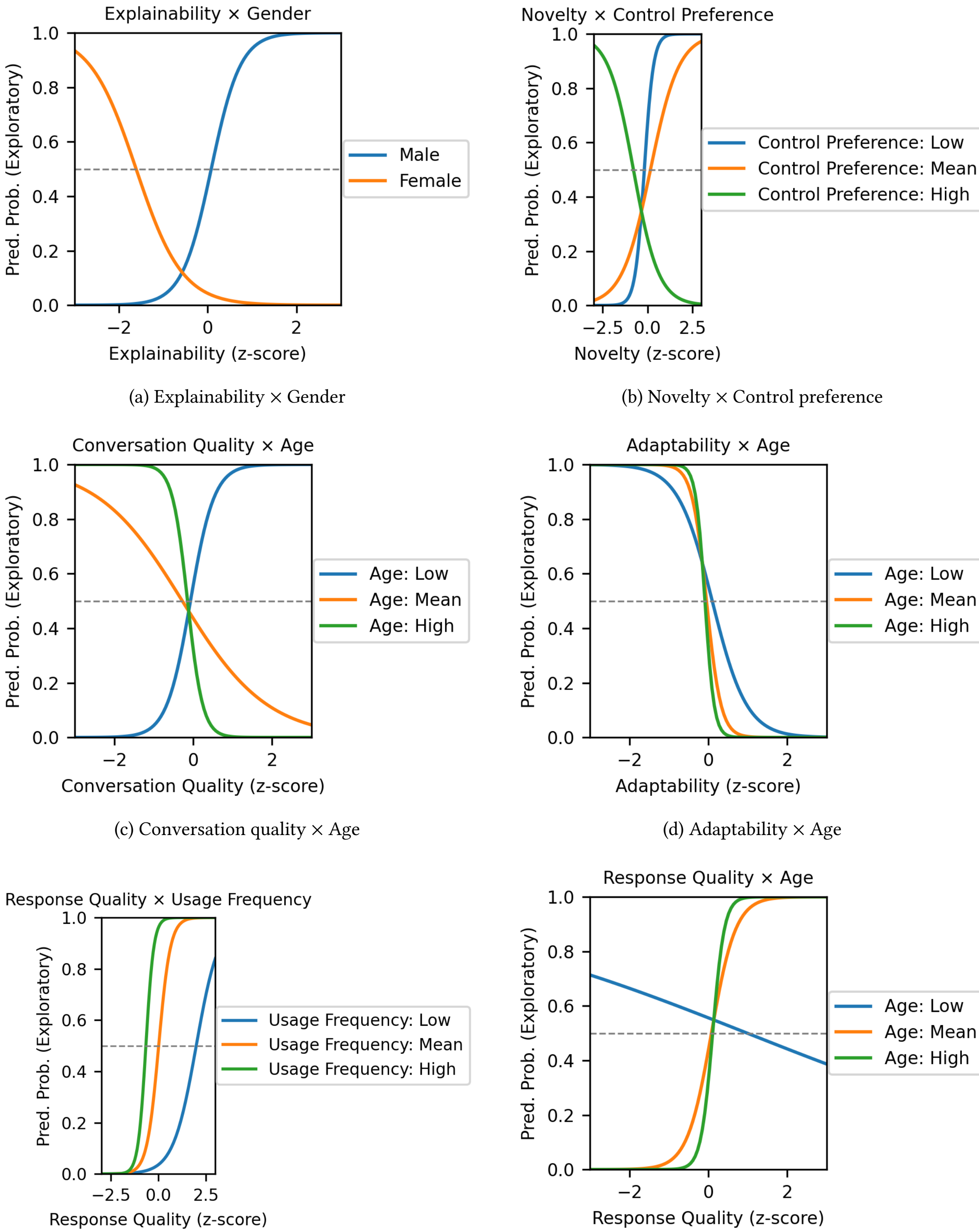
We summarise the outcomes of all nine preregistered hypotheses associated with RQ1 through RQ3. Table 7 presents each hypothesis alongside its empirical outcome and statistical evidence, offering a concise overview of which predictions were supported, partially supported, or not confirmed. Notably, hypotheses concerning novelty (H1a), as well as user clustering (H2a–H2b), received strong empirical support. In contrast, predictions grounded solely in accuracy (H1b) were not supported. Moderation effects emerged for usage frequency, control preference, age, and gender (H3a–H3c), highlighting the contextual conditions under which system quality beliefs influence interaction preferences.

5 Discussion

The findings advance the understanding of how user-valued system qualities, affective experiences, and individual traits dynamically converge to shape dialogue choice in CRS. We organised these findings around our three research questions to draw out both theoretical contributions and practical recommendations.

5.1 Key Insights

First, echoing work on hedonic system attributes in e-commerce and multimedia recommendation [10, 41, 88, 91], novelty and conversational quality emerged as robust predictors of preference for the *explorative* dialogue. Functional attributes such as accuracy and response quality, by contrast, showed limited predictive value when participants evaluated the dialogue styles side by side. This reinforces findings that affective and experiential dimensions often outweigh pure performance metrics in AI interactions [6, 36, 45, 57, 92]. **Second**, beyond these experiential drivers, perceived usefulness also forecast selection of the *explorative* style. This suggests that participants equated usefulness with informational depth and clarity, features more characteristic of open-ended exchanges than terse, *task-oriented* ones, and underscores the role of dialogue richness in user sensemaking [18, 60, 71]. **Third**, our clustering uncovered five latent profiles based on quality priorities, each showing distinct style preferences. This validates value-based segmentation as a cold-start personalisation strategy when behavioural histories are sparse [29, 33, 38, 78]. **Fourth**, moderation tests revealed that age, gender, and control expectations shape how quality beliefs translate into style choices. For instance, conversational clarity carried more weight for older users, while explainability effects differed



(a) Explainability × Gender

(b) Novelty × Control preference

(c) Conversation quality × Age

(d) Adaptability × Age

(e) Response quality × Usage frequency

(f) Response quality × Age

Figure 1: Significant moderation effects on preference for the *exploratory* interaction style. Each panel depicts the predicted probability of preferring exploratory interaction as a function of a standardised CRS quality score (z-score), stratified by levels of the moderator. See Table 6 for statistical details.

Table 7: Summary of hypothesis outcomes and supporting statistical evidence for RQ1–RQ3. *Note.* Odds ratios (OR) greater than 1 indicate increased likelihood of preferring the *exploratory* interaction. Statistical significance is denoted at $p \leq 0.05$.

Hypothesis	Outcome	Evidence
H1a. Preference for novelty or adaptability predicts exploratory interaction.	Partially supported	Novelty: OR = 1.66, p = .033; Adaptability: OR = .79, p = .366
H1b. Preference for accuracy and response quality predicts task-oriented interaction.	Not supported	Accuracy: p = .674; Response quality: p = .383
H1c. Enjoyment and surprise predict exploratory preference.	Partially supported	Enjoyment: OR = 2.19, p = .001; Surprise: p = .119
H1d. Usefulness and perceived effectiveness predict task-oriented preference.	Refuted	Usefulness: OR = 2.23, p = .003; Perceived effectiveness: OR = 2.17, p = .004
H2a. Users can be clustered by preferred CRS qualities.	Supported	k = 5 solution: silhouette = 0.146; ARI = 0.770 (SD = 0.214); elbow inflection at k = 5
H2b. Clusters differ in interaction preference.	Supported	$\chi^2(4, N = 139) = 10.47$, p = 0.033; Kruskal–Wallis H = 10.40, p = 0.034
H3a. CRS usage frequency moderates CRS quality×preference.	Supported	Response Quality × Usage Frequency: p = .041
H3b. Control preference moderates CRS quality×preference.	Supported	Novelty × Control Preference: p = .033
H3c. Age and gender moderate CRS quality×preference.	Supported	Explainability × Gender: p = .019; Conversation Quality × Age: p = .044; Adaptability × Age: p = .050; Response Quality × Age: p = .016

by gender. These nuances argue for adaptive systems that tailor both content and autonomy levels to user traits [16, 43, 51, 72, 84]. These insights pave the way for the theoretical elaborations and concrete design patterns that follow.

5.2 Theoretical and Design Implications

The findings contribute to theory by reframing interaction preference as a predictive construct influenced by system quality beliefs, retrospective UX appraisals, and individual differences. Building on the original CRS-Que framework [35], which treated system qualities solely as evaluation metrics, we demonstrate their forward-looking power in forecasting dialogue choice. In particular, novelty, conversational quality, and perceived usefulness, dimensions identified as user-valued, emerged as significant predictors of preference for *exploratory* interaction. This reinforces HCI theories that emphasise the formative role of experiential qualities in shaping expectations and behaviours through cognitive sensemaking and affective engagement [60, 68, 75, 86].

Unexpectedly, hypothesis H1d, which predicted that higher ratings of usefulness and perceived effectiveness would favour the *task-oriented* mode, was refuted. Instead, participants construed these instrumental dimensions in terms of informational completeness and clarity, aligning them more closely with *exploratory* dialogues than with terse exchanges. Consequently, *exploratory* interactions may be perceived as more "effective" because they provide essential context and explanations for comprehension and decision-making. This insight highlights the inherently dual cognitive and affective character of usefulness and perceived effectiveness, and suggests that genuine efficiency requires design innovations such as adaptive summarisation or information-density controls. These mechanisms preserve perceived utility while streamlining dialogue. Taken together, these theoretical advances point to a refined model of CRS adaptation, in which both experiential and instrumental dimensions jointly inform dynamic dialogue strategies.

Complementing these contributions, our identification of five latent user profiles based on quality-preference clustering offers empirical support for value-based segmentation in cold-start scenarios [15, 37, 38]. Each profile exhibited distinctive interaction preferences, validating segmentation heuristics that anticipate user expectations from stated priorities alone. Moderation analyses further extend autonomy-centred design paradigms by revealing how trait-level differences, such as age, gender, and control expectations, shape the weight users assign to specific system qualities when choosing between interaction styles [7, 16, 42, 73]. From a practical standpoint, these insights prescribe a new class of adaptive dialogue management systems. Specifically, designers should implement threshold-based switching rules that monitor real-time indicators (e.g. novelty sensitivity, enjoyment, control preference) to determine when to transition between task-oriented and exploratory strategies. Such heuristics reconcile diverse user needs, enabling systems to modulate tone, elaboration, and initiative according to inferred user values and traits.

These theoretical and design implications chart a path toward next-generation CRS architectures that transcend traditional performance metrics such as accuracy, precision, and recall. They integrate predictive preference modelling, robust value-profile segmentation, moderation-guided personalisation, and dynamic adaptation heuristics. In doing so, they lay the groundwork for conversational experiences that are truly user-centred, context-sensitive, and capable of driving higher satisfaction, loyalty, retention, and business performance.

5.3 Limitations and Future Work

Despite its contributions, this study has several limitations that point to specific next steps. First, the final sample size (N = 139) was adequate for the primary regression and clustering analyses. However, future research should recruit larger, more heterogeneous cohorts, including cross-cultural and domain-specific populations, to enable finer-grained moderation analyses and broader generalisability. Second, the use of scripted dialogues provided experimental control but abstracted from live CRS dynamics. To address this, our forthcoming prototype study will embed the threshold-based adaptive heuristics described above into a working conversational system. This implementation will support real-time system–user co-construction of recommendations and allow us to evaluate effects on user satisfaction and behavioural intentions (for example, continued use, purchase intent, or recommendation to others). Finally, we examined two distinct dialogue variants within a single domain. While this scope enabled rigorous hypothesis testing, subsequent work should extend this paradigm to multiple application areas and explore additional interaction styles, such as mixed-initiative or summarisation-driven flows. This would help assess the robustness and scalability of our adaptive strategies across longer-term engagements.

6 Conclusion

This study advances conversational recommender systems by moving beyond retrospective evaluation toward a predictive, adaptive framework that integrates system qualities, user experience appraisals, and individual-difference traits. We show that novelty, conversational quality, and usefulness reliably predict user preference for *task-oriented* or *exploratory* interaction styles, extending the CRS-Que framework into preference modelling. Clustering revealed five latent user profiles, supporting value-based segmentation as a viable strategy for cold-start scenarios. Moderation analyses further showed that age, gender, and control preferences shape how users interpret system qualities. Notably, the findings challenge the assumption that usefulness and effectiveness inherently favour *task-oriented* dialogues. When framed as signals of clarity and richness, these attributes were more aligned with exploratory interactions that promote sensemaking and empowerment. This suggests that instrumental UX dimensions often blur cognitive and affective boundaries. Consequently, task-oriented designs should embed contextual scaffolds, such as adaptive summarisation or information-density controls, to help users act on recommendations effectively. Practically, our results offer design principles for the next generation of CRS interfaces powered by generative AI. Systems that detect real-time signals such as novelty-seeking, enjoyment, or control preference can fluidly switch between interaction modes, improving both short-term satisfaction and long-term engagement. While grounded in recommendation, the framework also applies to tutoring, customer support, and personal assistant domains, where adapting to users' cognitive and emotional states can improve outcomes. This work contributes theoretical insights, adaptive design patterns, and a foundation for real-world evaluation. It sets the stage for conversational agents that do more than recommend. These systems can attune to user values, expectations, and emotional cues to deliver experiences that are context-sensitive, transparent, and genuinely responsive to human needs.

Acknowledgments

During the preparation of this work the author(s) used ChatGPT 4o in order to improve readability and language of the final draft. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

[1] Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M Pérez, and Iñigo Perona. 2013. An extensive comparative study of cluster validity indices. *Pattern recognition* 46, 1 (2013), 243–256.
[2] Shlomo Berkovsky, Jill Freyne, and Harri Oinas-Kukkonen. 2012. Influencing individually: fusing personalization and persuasion. 8 pages.
[3] Shlomo Berkovsky, Maurits C. Kaptein, and Massimo Zancanaro. 2016. Adaptivity and Personalization in Persuasive Technologies. In *Proceedings of the International Workshop on Personalization in Persuasive Technology (PPT 2016) (CEUR Workshop Proceedings, Vol. 1581)*, Rita Orji, Marc Reisinger, Markus Busch, Arie Dijkstra, Andre Stibe, and Manfred Tscheligi (Eds.). CEUR-WS.org, Salzburg, Austria, 13–25. https://hdl.handle.net/2066/161893 Held in conjunction with Persuasive Technology 2016.
[4] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2008. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction* 18 (2008), 245–286.
[5] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. *Advances in natural multimodal dialogue systems* (2005), 23–54.
[6] Ghazala Bilquise, Samar Ibrahim, and Khaled Shaalan. 2022. Emotionally intelligent chatbots: A systematic literature review. *Human Behavior and Emerging Technologies* 2022, 1 (2022), 9601630.
[7] Sofia Bonicalzi, Mario De Caro, and Benedetta Giovanola. 2023. Artificial intelligence and autonomy: on the ethical dimension of recommender systems. *Topoi* 42, 3 (2023), 819–832.
[8] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation* 42 (2008), 335–359.
[9] James Carifio and Rocco Perla. 2008. Resolving the 50-year debate around using and misusing Likert scales. 1150–1152 pages.
[10] Pablo Castells, Neil Hurley, and Saul Vargas. 2021. Novelty and diversity in recommender systems. In *Recommender systems handbook*. Springer, 603–646.
[11] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758.
[12] Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction* 22 (2012), 125–150.
[13] Evangelia Christakopoulou and Katja Hofmann. 2016. Towards Conversational Recommender Systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 815–824. doi:10.1145/2939672.2939746
[14] Jacob Cohen. 1992. A power primer. *Psychological bulletin* 112, 1 (1992), 155.
[15] Alvise De Biasio, Andrea Montagna, Fabio Aiolli, and Nicolò Navarin. 2023. A systematic review of value-aware recommender systems. *Expert Systems with Applications* 226 (2023), 120131.
[16] Juan Ignacio del Valle and Francisco Lara. 2024. AI-powered recommender systems and the preservation of personal autonomy. *AI & SOCIETY* 39, 5 (2024), 2479–2491.
[17] Yashar Deldjoo, Z He, J McAuley, A Korikov, S Sanner, A Ramisa, R Vidal, M Sathiamoorthy, A Kasirzadeh, and S Milano. 2024. A Review of Modern Recommender Systems Using Generative Models (Gen-RecSys). ArXiv. 2024. *arXiv preprint arXiv:2404.00579* (2024).
[18] Brenda Dervin. 1998. Sense-making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of knowledge management* 2, 2 (1998), 36–46.
[19] Hongkai Du, Haozhou Li, Qinke Peng, and Laiyi Fu. 2024. User Linguistic Style Awareness and Interest-Driven Conversational Recommender Systems. In *2024 IEEE 4th International Conference on Digital Twins and Parallel Intelligence (DTPI)*. IEEE, 215–220.
[20] Siamak Farshidi, Kiyan Rezaee, Sara Mazaheri, Amir Hossein Rahimi, Ali Dadashzadeh, Morteza Ziabakhsh, Sadegh Eskandari, and Slinger Jansen. 2024. Understanding user intent modeling for conversational recommender systems: a systematic literature review. *User Modeling and User-Adapted Interaction* (2024), 1–64.
[21] Andy Field. 2024. *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
[22] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience* 5, 1 (2020), 3.
[23] Jodi Forlizzi and Katja Battarbee. 2004. Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. 261–268.
[24] Chris Fraley and Adrian E Raftery. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal* 41, 8 (1998), 578–588.
[25] Krzysztof Z Gajos. 2008. *Automatically generating personalized user interfaces*. University of Washington.
[26] Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten De Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems:

A survey. *AI open* 2 (2021), 100–126.
[27] Shirley Gregor and Izak Benbasat. 1999. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly* (1999), 497–530.
[28] Marc Hassenzahl. 2010. *Experience design: Technology for all the right reasons*. Vol. 8. Morgan & Claypool Publishers.
[29] Shunichi Hattori and Yasufumi Takama. 2014. Recommender system employing personal-value-based user model. *Journal of advanced computational intelligence and intelligent informatics* 18, 2 (2014), 157–165.
[30] Kasper Hornbæk and Morten Hertzum. 2017. Technology acceptance and user experience: A review of the experiential component in HCI. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 5 (2017), 1–30.
[31] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*. John Wiley & Sons.
[32] Dietmar Jannach. 2023. Evaluating conversational recommender systems: A landscape of research. *Artificial Intelligence Review* 56, 3 (2023), 2365–2400.
[33] Dietmar Jannach and Himan Abdollahpouri. 2023. A survey on multi-objective recommender systems. *Frontiers in big Data* 6 (2023), 1157899.
[34] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)* 54, 5 (2021), 1–36.
[35] Yucheng Jin, Li Chen, Wanling Cai, and Xianglin Zhao. 2024. CRS-Que: A user-centric evaluation framework for conversational recommender systems. *ACM Transactions on Recommender Systems* 2, 1 (2024), 1–34.
[36] Jeong Soo Kim, Minseong Kim, and Tae Hyun Baek. 2025. Enhancing user experience with a generative AI chatbot. *International Journal of Human–Computer Interaction* 41, 1 (2025), 651–663.
[37] Bart P Knijnenburg and Martijn C Willemsen. 2015. Evaluating recommender systems with user experiments. In *Recommender systems handbook*. Springer, 309–352.
[38] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504.
[39] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: systematic review. *Journal of medical Internet research* 21, 11 (2019), e15360.
[40] Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction* 22 (2012), 101–123.
[41] Philip Kossack and Herwig Unger. 2023. Emotion-aware Chatbots: understanding, reacting and adapting to human emotions in text conversations. In *International Conference on Autonomous Systems*. Springer, 158–175.
[42] Joshua Krook and Jan Blockx. 2023. Recommender systems, autonomy and user engagement. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*. 1–9.
[43] Guy Laban and Theo Araujo. 2020. The effect of personalization techniques in users' perceptions of conversational recommender systems. In *Proceedings of the 20th ACM international conference on intelligent virtual agents*. 1–3.
[44] Kleanthi Lakiotaki, Nikolaos F Matsatsinis, and Alexis Tsoukias. 2011. Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems* 26, 2 (2011), 64–76.
[45] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2022. *Measuring user engagement*. Springer Nature.
[46] Nuno Laranjeiro, Seyma Nur Soydemir, and Jorge Bernardino. 2015. A survey on data quality: classifying poor data. In *2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC)*. IEEE, 179–188.
[47] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold POS Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 719–728.
[48] Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. Interactive path reasoning on graph for conversational recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2073–2083.
[49] Lei Li, Yongfeng Zhang, Dugang Liu, and Li Chen. 2023. Large language models for generative recommendation: A survey and visionary discussions. *arXiv preprint arXiv:2309.01157* (2023).
[50] Yuan Ma and Jürgen Ziegler. 2023. Initiative transfer in conversational recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 978–984.
[51] Yuan Ma and Jürgen Ziegler. 2024. Investigating meta-intents: user interaction preferences in conversational recommender systems. *User Modeling and User-Adapted Interaction* 34, 5 (2024), 1535–1580.
[52] Tariq Mahmood and Francesco Ricci. 2009. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*. 73–82.
[53] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.
[54] John McCarthy and Peter Wright. 2004. Technology as experience. *interactions* 11, 5 (2004), 42–43.
[55] Michael Frederick McTear, Zoraida Callejas, and David Griol. 2016. *The conversational interface*. Vol. 6. Springer.
[56] Adam W Meade and S Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological methods* 17, 3 (2012), 437.
[57] MIT Media Lab and OpenAI. 2025. Affective Use and Emotional Wellbeing in ChatGPT: Insights from Collaborative Evaluation. https://www.media.mit.edu/posts/openai-mit-research-collaboration-affective-use-and-emotional-wellbeing-in-ChatGPT/. Accessed: 2025-06-05.
[58] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15 (2010), 625–632.
[59] Heather L O'Brien and Elaine G Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.
[60] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.
[61] Peter Pirolli and Daniel M Russell. 2011. Introduction to this special issue on sensemaking. 8 pages.
[62] Alina Pommeranz, Joost Broekens, Pascal Wiggers, Willem-Paul Brinkman, and Catholijn M Jonker. 2012. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. *User Modeling and User-Adapted Interaction* 22 (2012), 357–397.
[63] Dhanya Pramod and Prafulla Bafna. 2022. Conversational recommender systems techniques, tools, acceptance, and adoption: a state of the art review. *Expert Systems with Applications* 203 (2022), 117539.
[64] Bilih Priyogi. 2019. Preference elicitation strategy for conversational recommender system. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 824–825.
[65] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.
[66] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 22 (2012), 317–355.
[67] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2024. User modeling and user profiling: A comprehensive survey. *arXiv preprint arXiv:2402.09660* (2024).
[68] Arpit Rana, Scott Sanner, Mohamed Reda Bouadjenek, Ronald Di Carlantonio, and Gary Farmaner. 2024. User experience and the role of personalization in critiquing-based conversational recommendation. *ACM Transactions on the Web* 18, 4 (2024), 1–21.
[69] Verena Rieser and Oliver Lemon. 2011. *Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation*. Springer Science & Business Media.
[70] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 269–276.
[71] Reijo Savolainen. 1993. The sense-making theory: Reviewing the interests of a user-centered approach to information seeking and use. *Information processing & management* 29, 1 (1993), 13–28.
[72] Max Schemmer, Andrea Bartos, Philipp Spitzer, Patrick Hemmer, Niklas Kühl, Jonas Liebschner, and Gerhard Satzger. 2023. Towards effective human-ai decision-making: The role of human learning in appropriate reliance on ai advice. *arXiv preprint arXiv:2310.02108* (2023).
[73] Rahul Singh. 2024. The Recommender System Paradox: Autonomy in the Age of AI Recommendations. In *2024 3rd International Joint Conference on Information and Communication Engineering (JCICE)*. IEEE, 139–143.
[74] Clemencia Siro, Mohammad Aliannejadi, and Maarten De Rijke. 2023. Understanding and predicting user satisfaction with conversational recommender systems. *ACM Transactions on Information Systems* 42, 2 (2023), 1–37.
[75] Åsne Stige, Efpraxia D Zamani, Patrick Mikalef, and Yuzhen Zhu. 2024. Artificial intelligence (AI) for user experience (UX) design: a systematic literature review and future research agenda. *Information Technology & People* 37, 6 (2024), 2324–2352.
[76] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *The 41st international acm sigir conference on research & development in information retrieval*. 235–244.
[77] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Modeling and User-Adapted Interaction* 22 (2012), 399–439.
[78] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*. Springer, 353–382.

[79] Chun-Hua Tsai and Peter Brusilovsky. 2021. The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction* 31 (2021), 591–627.

[80] Svitlana Vakulenko, Ilya Markov, and Maarten de Rijke. 2017. Conversational exploratory search via interactive storytelling. *arXiv preprint arXiv:1709.05298* (2017).

[81] Ronald L Wasserstein and Nicole A Lazar. 2016. The ASA statement on p-values: context, process, and purpose. 129–133 pages.

[82] Ryen W White and Resa A Roth. 2009. *Exploratory search: Beyond the query-response paradigm.* Number 3. Morgan & Claypool Publishers.

[83] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, et al. 2024. A survey on large language models for recommendation. *World Wide Web* 27, 5 (2024), 60.

[84] Mengke Wu, Weizi Liu, Yanyun Wang, and Mike Yao. 2025. Negotiating the Shared Agency between Humans & AI in the Recommender System. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[85] Wei Xu, Jue Xiao, and Jianlong Chen. 2024. Leveraging large language models to enhance personalized recommendations in e-commerce. In *2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE)*. IEEE, 1–6.

[86] Yang Xu, Yingchia Liu, Haosen Xu, and Hao Tan. 2024. AI-driven UX/UI design: Empirical research and applications in FinTech. *Academia Nexus Journal* 3, 1 (2024).

[87] Sojeong Yun and Youn-kyung Lim. 2025. User Experience with LLM-powered Conversational Recommendation Systems: A Case of Music Recommendation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–15.

[88] Yizhe Zhang, Yucheng Jin, Li Chen, and Ting Yang. 2024. Navigating user experience of chatgpt-based conversational recommender systems: The effects of prompt guidance and recommendation domain. *arXiv preprint arXiv:2405.13560* (2024).

[89] Zheng Zhang, Ryuichi Takanobu, Qi Zhu, MinLie Huang, and XiaoYan Zhu. 2020. Recent advances and challenges in task-oriented dialog systems. *Science China Technological Sciences* 63, 10 (2020), 2011–2027.

[90] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, et al. 2024. Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering* (2024).

[91] Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1006–1014.

[92] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics* 46, 1 (2020), 53–93.

A Appendix: Scripted Dialogue Excerpts

This appendix provides representative excerpts from the two scripted dialogues used as stimuli in the study, illustrating the key differences between the *task-oriented* and *exploratory* interaction described in Section 2.1 and operationalised in Section 3.1.

A.1 Task-oriented Interaction

System: Welcome! What type of dress are you looking for today?
User: I want a casual dress.
System: Great! Here are the top 3 casual dresses we recommend. The first one is priced at $59.99 with a 10% discount.
User: I like the first one.
System: Excellent choice! Would you like to add it to your cart?

This dialogue illustrates the concise and goal-directed nature of the *task-oriented* interaction, where the system quickly narrows down options and facilitates rapid decision-making with minimal elaboration.

A.2 Exploratory Interaction

System: Hello! Looking for a dress today? What styles do you usually enjoy?
User: I like casual dresses mostly.
System: That's wonderful! Casual dresses can range widely from flowy maxi dresses to chic shirt dresses. Are you interested in exploring some trendy new arrivals or classic styles?
User: Maybe something trendy.
System: Perfect! How about a light floral maxi dress? It's popular this season and perfect for both casual outings and special occasions. I can also suggest some accessories to complement the look.
User: Sounds interesting! Show me more options.
System: Certainly! Here are some other trendy choices along with styling tips.

This excerpt demonstrates the *exploratory* interaction's open-ended prompts, elaborative responses, and reflective dialogue, designed to encourage discovery and deliberation.

B Appendix: Clustering Diagnostics

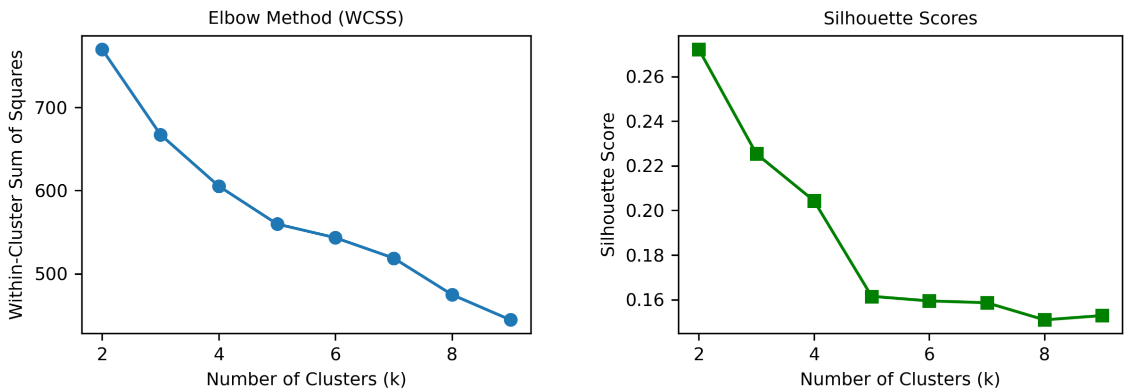
We demonstrated that a five-cluster solution most effectively differentiates participants by their CRS quality importance ratings (Section 4.2). Figure A.1 presents the elbow plot (within-cluster sum of squares) and silhouette scores for $k = 2$ through 9. These diagnostics, together with the stability and preference-association metrics in Table A.1, confirm that $k = 5$ achieves an optimal compromise between cluster cohesion and separation.

Table A.1: Evaluation metrics for hierarchical (Ward's method) clustering solutions ($k = 3$–9, $N = 139$). "Min/Max" indicates cluster size range. Silhouette is the average silhouette coefficient. ARI (mean, SD) assesses cluster stability via bootstrap. χ^2 and H test associations with interaction preference.

k	Min/Max	Silh.	Mean	SD	χ^2	p_{χ^2}	H	p_H
3	18/87	0.182	0.730	0.287	14.92	0.0006	14.81	0.0006
4	18/63	0.150	0.747	0.203	14.93	0.0019	14.82	0.0020
5	18/37	0.146	0.770	0.214	14.93	0.0048	14.82	0.0051
6	3/37	0.153	0.714	0.185	15.34	0.0090	15.23	0.0094
7	3/34	0.127	0.731	0.185	15.38	0.0175	15.27	0.0182
8	3/34	0.131	0.774	0.190	15.38	0.0314	15.27	0.0327
9	3/26	0.117	0.777	0.168	16.03	0.0420	15.91	0.0436

C Appendix: Cluster Profiles

As detailed in Section 4.2, we applied hierarchical clustering using Ward's method to z-standardised importance ratings of eight CRS qualities from $N = 139$ participants. Figure A.2 visualises the five-cluster solution. Each cell reflects the mean importance rating for a given quality—accuracy, explainability, novelty, conversational quality, attentiveness, understanding, adaptability, and response quality—within each user cluster.



(a) Elbow plot

(b) Silhouette scores

Figure A.1: Clustering diagnostics supporting the five-cluster solution for RQ2. (a) Elbow plot of within-cluster sum of squares (WCSS) for $k = 2$ through 9; the inflection at $k = 5$ suggests diminishing returns beyond this point. (b) Silhouette scores plateau near $k = 5$, indicating a balanced trade-off between cluster cohesion and separation.

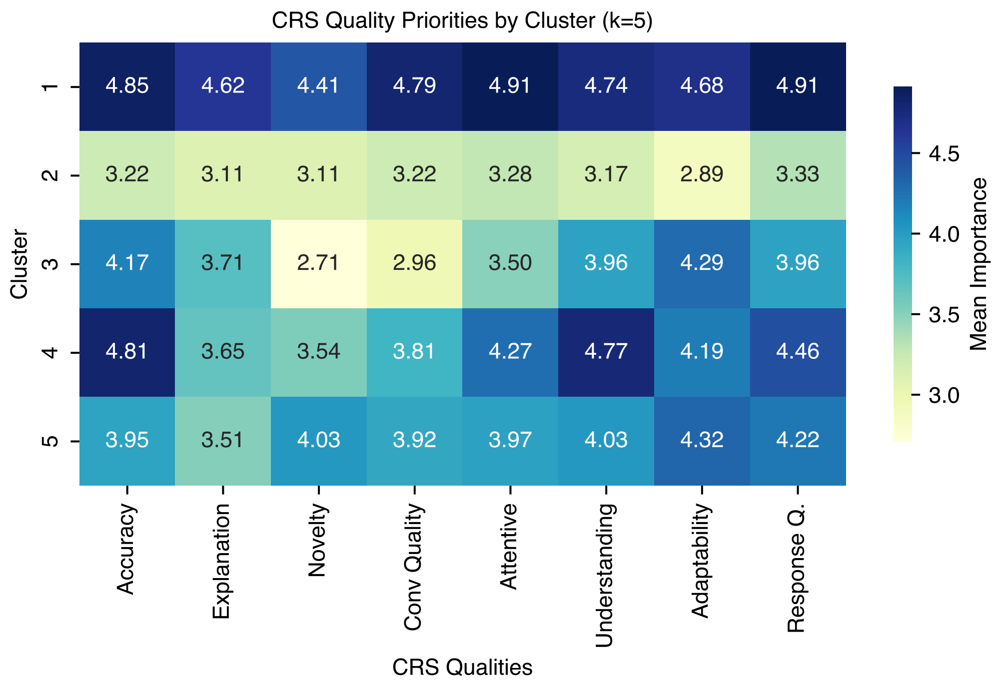


Figure A.2: Heatmap of mean importance ratings (1 = low, 5 = high) across five user clusters derived via hierarchical clustering (Ward's method, $k = 5$, $N = 139$). Each row represents a CRS quality, each column a user cluster. Darker cells indicate stronger average importance ratings.