International Journal on Management Education & Emerging Technology
ISSN: 2583-9438

Understanding The Information-Seeking Behaviors And Credibility Evaluation Practices Of Graduate Researchers During Web-Based Scholarly Information Retrieval: A Phenomenological Study

Janepol G. Ballard[1], Reagan Ricafort[2]
Doctor in Information Technology, AMA University, Quezon City, Philippines
janepol.ballard@ustp.edu.ph, Reagan.ricafort@ama.edu.ph

Abstract

Web-based scholarly retrieval has become the primary route through which graduate researchers build knowledge, yet the abundance, heterogeneity, and machine-mediation of today's web make locating and trusting sources newly difficult. This study explores the lived experience of that work. Using a descriptive (Husserlian) phenomenological design, it examines how graduate researchers describe their information-seeking behaviors, how they evaluate source credibility, what meanings and tensions characterize the experience, and how generative artificial intelligence is reshaping both. Data will be generated through in-depth semi-structured interviews, a concurrent think-aloud retrieval task, and review of participant-supplied search artifacts, with a criterion-purposive sample analyzed through Colaizzi's seven-step method. This paper presents the Introduction and Methodology and the complete research instrument. It contributes a first-person, meaning-centered account intended to inform information-literacy instruction, academic search design, and doctoral supervision in the generative-AI era.

Article Info

Keywords: Information-Seeking Behavior, Credibility Evaluation, Web Information Retrieval, Graduate Researchers, Phenomenology.

Date of Submission: 11/06/.2026
Date of Review: 13/07/2026
Date of Acceptance: 20/07/2026
IJMEET / Volume 4, Issue 2, 2026
ISSN: 2583-9438

1. INTRODUCTION

1.1 Background and Context

The migration of scholarly communication from curated print repositories to networked, algorithmically mediated environments has fundamentally reconfigured how graduate researchers locate, appraise, and appropriate knowledge. Information seeking is no longer a discrete transaction concluded at a reference desk; it is a sustained, iterative, and affectively charged activity that now unfolds across general search engines, disciplinary databases, institutional repositories, preprint servers, and, most recently, conversational artificial intelligence. Foundational theory anticipated part of this complexity. Kuhlthau's Information Search Process (ISP) established that seeking is experienced in stages that couple cognitive work with shifting feelings, uncertainty and confusion early, clarity and confidence late, rather than as a purely rational retrieval act [1]. Wilson's successive models of information behaviour situated that process within the person's context, motivations, and the intervening variables that facilitate or obstruct access [2], while Ellis's behavioural characterisation supplied a granular vocabulary, starting, chaining, browsing, differentiating, monitoring, and extracting, for the micro-moves a researcher performs while traversing a corpus [3].

These frameworks were, however, forged in comparatively stable and professionally curated information ecologies. The contemporary web is different in kind, not merely in degree. It is abundant to the point of saturation, epistemically heterogeneous, and in places adversarial: peer-reviewed scholarship now coexists with predatory journals, search-optimized promotional content, and fluent machine-generated text whose provenance is difficult to establish. Where an earlier researcher's central difficulty was access, reaching the material at all, the graduate researcher today more often confronts the inverse problem of discernment: deciding, amid a surfeit of plausible-looking results, which sources merit trust and citation. The locus of scholarly labour has, in effect, shifted from retrieval toward evaluation.

Two features of the present environment deserve emphasis because they bear directly on the experience under study. The first is algorithmic mediation: results are ranked, personalized, and filtered by systems whose criteria are opaque to the researcher, so that the very order in which knowledge presents itself is shaped by processes the user neither sees nor controls. Where the card catalogue once offered a stable, legible ordering of a bounded collection, the search engine offers a fluid, personalized ordering of an effectively unbounded one, convenient, but epistemically consequential in ways that are easy to overlook. The second is the erosion of the boundary between the scholarly and the open web. A single results page may interleave a peer-reviewed article, a preprint of uncertain status, an institutional press release, a commercial explainer, and, increasingly, an AI-generated summary, each competing for attention and trust. The graduate researcher must therefore perform, often tacitly and under time pressure, an ongoing triage that earlier generations could largely delegate to the gatekeeping of librarians and publishers. It is this quietly demanding, largely invisible evaluative work, rather than the mechanics of querying, that the present study takes as its phenomenon.

The graduate researcher is a revealing case for studying this work. Unlike the casual searcher, the graduate researcher seeks not merely to be informed but to build a citable, defensible argument; a source is retained not only because it answers a question but because it can withstand the scrutiny of examiners, reviewers, and disciplinary peers. Citation, in this sense, is an act of scholarly identity and accountability as much as of information use. The stakes, a dissertation, a qualifying examination, a first publication, make the consequences of a misjudged source materially and reputationally real. For this population, then, seeking and appraisal are not incidental skills but constitutive of the work of becoming a scholar, which is precisely why their lived experience of the process warrants description in its own right rather than measurement against an external rubric.

1.2 The Problem Situation and Its Evidence

A recurring and well-evidenced problem is that facility with technology does not translate into competence in judging it. Wineburg and McGrew's comparative study of professional fact-checkers, doctoral historians, and undergraduates found that the two academic groups tended to read "vertically", remaining within an unfamiliar page, scrutinizing its prose and internal cues, whereas fact-checkers read "laterally," leaving the page almost immediately to interrogate the source from the wider web, and reached better-warranted judgments in far less time [4]. That expert scholars themselves defaulted to weaker strategies is a sobering indication that disciplinary training does not automatically confer web-credibility skill. Complementary intervention research

shows the deficit is remediable: Breakstone and colleagues demonstrated that college students' evaluation of online sources improved significantly after a short asynchronous course modelling fact-checkers' strategies, which by implication confirms a weak baseline [5].

The difficulty is neither confined to one setting nor limited to novices. Studies of students' evaluation of online information across contexts consistently report that surface cues, design polish, familiar domain suffixes, and the sheer position of a result in a ranked list, exert disproportionate influence on trust, while deeper signals of provenance and method are underused [6], [7]. Because such cues are cheap to fake and easy to read, they reward the confident but shallow appraisal that expert fact-checkers explicitly avoid [4]. The result is a persistent gap between the self-perceived competence of technologically fluent researchers and their demonstrated evaluative practice, a gap that ordinary confidence tends to conceal.

Why evaluation falters is partly explained by the cognitive economics of the web. Metzger and Flanagin argued that, faced with more information than can be systematically vetted, users routinely fall back on cognitive heuristics, reputation, endorsement, consistency, and the like, rather than effortful appraisal [6]. Instruments developed to measure the skill corroborate wide variability: the EVON screening tool, validated with university students, treats relevance and credibility judgement as distinct, learnable competencies that many students exercise unevenly [7]. Onto this already strained landscape has arrived generative artificial intelligence, which both eases and imperils scholarly search. Baek, Tate, and Warschauer found college students describing tools such as ChatGPT as almost "too good to be true," valuing their fluency while remaining wary of accuracy and dependence [8]; Chan and Hu, canvassing students' own voices, documented parallel enthusiasm and concern over reliability, transparency, and academic integrity [9]. The distinctive hazard is that generative systems can produce authoritative-sounding syntheses and even plausibly formatted but non-existent references, relocating the credibility problem from the ranking of retrieved documents to the trustworthiness of synthesized claims themselves.

1.3 The Research Gap

Despite a substantial and growing literature, three gaps persist. First, the dominant evidentiary mode is quantitative or task-based, scored assessments, surveys, and controlled experiments that measure how well participants perform against a rubric, rather than accounts of how the process is lived and made sense of from the inside. Second, most studies centre undergraduates and general internet users, leaving the graduate researcher comparatively under-examined even though this population faces higher stakes (dissertations, examinations, and publication), operates under discipline-specific norms of evidence, and is expected to model rigorous practice. Third, the swift diffusion of generative AI has outpaced qualitative understanding of how it is being woven into, and is unsettling, established seeking and appraisal routines. In sum, the field knows a good deal about whether users can evaluate sources and comparatively little about what the entire experience means to those for whom scholarly retrieval is a defining professional practice. A phenomenological inquiry is well suited to this gap, because its purpose is to describe the essence of a lived experience rather than to quantify a competence.

1.4 Research Objectives and Questions

The general objective of this study is to describe and interpret the essence of graduate researchers' lived experience of seeking and evaluating scholarly information on the web. Four questions guide the inquiry:

RQ1. How do graduate researchers describe their information-seeking behaviors during web-based scholarly retrieval?

RQ2. How do graduate researchers evaluate the credibility of the sources they encounter during this process?

RQ3. What meanings, emotions, and tensions characterize their lived experience of scholarly retrieval and appraisal?

RQ4. How are these seeking behaviors and credibility practices being reshaped by generative artificial intelligence tools?

1.5 Significance of the Study

The study is significant on practical, methodological, and theoretical grounds. Practically, a fine-grained description of how graduate researchers actually seek and judge information can inform information-literacy

curricula, moving instruction beyond checklist heuristics toward the lateral, source-interrogating dispositions that distinguish expert evaluators [4], [5]. For designers of academic search systems and library discovery interfaces, first-person insight into where trust is granted or withheld can guide provenance cues and transparency features, an increasingly urgent design concern as retrieval becomes conversational and synthetic. For doctoral supervisors and graduate programs, the findings offer a vocabulary for mentoring students through the affective and evaluative difficulties of the process. Methodologically, the study contributes a phenomenological lens to a domain long dominated by experimental measurement. Theoretically, it extends process- and behaviour-oriented models of information seeking [1]–[3] and heuristic accounts of credibility [6] into the generative-AI era, testing their descriptive adequacy against contemporary lived experience.

2. METHODOLOGY

2.1 Research Design

This study adopts a qualitative, descriptive phenomenological design grounded in the Husserlian tradition and operationalized through Moustakas's procedures for transcendental phenomenological research [10]. Phenomenology is chosen because the research questions ask not how frequently or how well graduate researchers behave, but what their seeking and appraisal actually mean as lived, the intentional structure of the experience and its essential, invariant features. Descriptive phenomenology is preferred over its interpretive (hermeneutic) counterpart because the aim is to render the phenomenon as participants experience it, with the researcher's pre-understandings deliberately set aside through epoché, or bracketing, rather than mobilized as an interpretive resource. The design is likewise distinguished from adjacent qualitative approaches: unlike grounded theory, the study does not seek to build a predictive theory of causes; unlike case study, it is bounded by a shared experience rather than by a site or organization; and unlike generic thematic inquiry, it is disciplined by phenomenological philosophy at every stage, from bracketing to the derivation of an exhaustive description [12]. Bracketing is treated as an ongoing practice: the researcher records assumptions about "good" searching and credible sources in a reflexive journal before and throughout fieldwork, so that these do not silently structure either data generation or analysis.

Three commitments of the descriptive tradition organize the design. The first is intentionality: consciousness is always consciousness of something, so the study attends not to searching in the abstract but to searching-as-directed-toward trustworthy scholarship, taking the participant's relation to the object of experience as primary. The second is phenomenological reduction, the movement from the natural attitude, in which the everydayness of search is taken for granted, to a reflective stance in which the familiar can appear as if encountered for the first time. The third is the eidetic aim of moving beyond individual idiosyncrasy toward the essential structure that any instance of the experience must possess. These commitments are not ornamental; they discipline concrete choices, from the wording of experience-near questions to the refusal, in analysis, to substitute the researcher's categories for the participant's described world.

2.2 Data Sources

The study draws on primary and secondary sources. The primary sources are the first-person accounts of graduate researchers, their narrated recollections in interview and their concurrent verbalizations during a live retrieval task, since these carry the experiential meaning the study seeks. Secondary sources are participant-supplied artifacts that contextualize and triangulate these accounts: browser search histories or query logs, saved reference lists and citation-manager libraries, and screenshots of results the participant chose to keep or reject. Pertinent institutional documents, such as library research guides and program-level policies on the use of AI, are consulted only as background to situate the experience, not as freestanding units of analysis. Artifacts are used to illuminate and cross-check narrative claims, addressing the familiar gap between what people say they do and what they are observed to do, and never to override the participant's own account of meaning.

2.3 Data Collection Methods

Three complementary techniques are used in a single, sequenced session per participant. First, an in-depth semi-structured interview elicits retrospective accounts of habitual seeking and credibility practices, structured by the guide presented in Section 3 but responsive to each participant's narrative. Second, a concurrent think-aloud web-retrieval task invites the participant to perform a realistic scholarly search while verbalizing their

reasoning, allowing the phenomenon to be observed as enacted rather than only as recalled; think-aloud is well matched to phenomenology because it surfaces in-the-moment intentionality and the tacit cues on which trust turns. Third, a brief document and log review is conducted collaboratively at the close of the session, with the participant narrating the artifacts they have chosen to share. Interviews are expected to run approximately sixty to ninety minutes, are audio-recorded with consent, and, for the think-aloud segment, accompanied by optional screen capture. All recordings are transcribed verbatim, with pauses and hesitations preserved, because such features often carry phenomenological significance for the experience of uncertainty [14].

Concurrent think-aloud is chosen deliberately over retrospective report, because the study's interest lies precisely in the in-the-moment intentionality of appraisal, the fleeting cues, hesitations, and micro-judgments that participants seldom recall accurately once a search is over. To reduce the reactivity that observation can introduce, each participant first completes a brief, low-stakes warm-up search so that verbalizing becomes habitual before the recorded task begins, and the facilitator intervenes only minimally and non-directively, prompting to sustain articulation rather than to steer the search. In this way the three techniques triangulate the same phenomenon from complementary angles: the interview recovers its remembered meaning, the think-aloud exposes its enacted texture, and the artifacts anchor both to concrete traces of practice.

2.4 Sampling Approach

Participants are selected through criterion-based purposive sampling, the standard strategy in phenomenology, so that every participant has substantial, direct experience of the phenomenon under study. The target sample is twelve to fifteen graduate researchers, a range consistent with the depth of phenomenological analysis; the final number is governed by the principle of data saturation, understood here as information power, data collection continues until additional accounts cease to yield new constituents of the experience. Inclusion and exclusion criteria are summarized in Table 2.

Table 2. Participant inclusion and exclusion criteria.

Dimension	Inclusion criteria	Exclusion criteria
Enrolment	Currently enrolled in a master's or doctoral program.	Undergraduate or non-degree students.
Research activity	Actively conducting a thesis, dissertation, or supervised research requiring literature retrieval.	No current requirement to retrieve scholarly literature.
Web experience	Routinely uses the web and academic databases for scholarly search (≥ 6 months).	Relies exclusively on materials supplied by others.
AI exposure	Any level of experience with, or deliberate avoidance of, generative AI tools (both are informative).	,
Consent & language	Provides informed consent and can articulate experiences in the interview language.	Unable or unwilling to give informed consent.

Recruitment proceeds through graduate program coordinators and disciplinary mailing lists, supplemented by limited snowball referral to reach information-rich cases across fields, since disciplinary culture is likely to inflect both seeking and credibility norms.

Sample size is treated as a matter of information power rather than a fixed quota: the richer and more on-point the accounts, and the more focused the analytic aim, the fewer participants are required to render the phenomenon's structure. Sampling also attends to maximum variation within the criterion group, across disciplines, stages of candidature, and levels of generative-AI adoption, so that the essence derived is not an artifact of one subfield's conventions but holds across meaningfully different vantage points on the same

experience. A participant who deliberately avoids AI tools is, on this logic, as informative as an enthusiastic adopter, since both illuminate the phenomenon from a distinct angle.

2.5 Analytical Framework

Data are analyzed using Colaizzi's seven-step method of descriptive phenomenological analysis [11], complemented by the phenomenologically grounded approach to thematic analysis articulated by Sundler and colleagues [12]. The steps are applied iteratively: (1) each transcript is read repeatedly to acquire a holistic sense of the account; (2) significant statements pertaining directly to seeking and credibility are extracted; (3) meanings are formulated from these statements while remaining faithful to the participants' words; (4) formulated meanings are organized into clusters of themes; (5) findings are integrated into an exhaustive description of the phenomenon; (6) that description is condensed into a statement of its fundamental structure, its essence; and (7) the structure is returned to participants for validation. Throughout, the theoretical models introduced earlier function as sensitizing lenses rather than coding templates: the process stages of the ISP [1], the contextual variables of Wilson's model [2], Ellis's behavioural moves [3], and the heuristic account of credibility [6] inform the researcher's attentiveness without predetermining the themes, which are allowed to emerge from the data. The alignment of research questions, interview domains, data sources, and sensitizing lenses is shown in Table 1, and the overall procedural flow of the inquiry is depicted in Figure 1. Coding and retrieval are managed in qualitative data-analysis software to maintain an organized, auditable corpus.

A worked illustration clarifies the analytic movement. A participant's remark, "I saw it was from a journal I'd never heard of, so I opened a new tab to see who publishes it", is extracted as a significant statement; its formulated meaning might be rendered as "unfamiliarity of the venue triggers external verification of the source's provenance"; and, aggregated with kindred meanings, it may contribute to a theme such as "trust as something checked beyond the page." The essence sought at the sixth step is not any particular tactic but the invariant structure that makes the experience what it is across participants, what, at bottom, it is like to seek scholarship one can defensibly cite. This movement from concrete utterance to essential structure is the disciplined core of descriptive phenomenology and the safeguard against reducing rich accounts to a mere inventory of behaviours.

Table 1. Alignment of research questions, interview domains, data sources, and sensitizing lenses.

RQ	Interview domain	Primary data source	Sensitizing lens
RQ1	Habitual seeking routines, tools, query behaviour, navigation moves.	Interview; think-aloud task; query logs.	ISP [1]; Wilson [2]; Ellis [3].
RQ2	Cues and reasoning used to judge source trustworthiness.	Think-aloud task; interview; kept/rejected results.	Cognitive heuristics [6]; lateral reading [4].
RQ3	Emotions, uncertainty, confidence, tensions, meaning.	Interview; think-aloud verbalizations.	Affective stages of the ISP [1].
RQ4	Use, distrust, and integration of generative AI in search.	Interview; artifacts; observed tool use.	Student GenAI experience [8], [9].

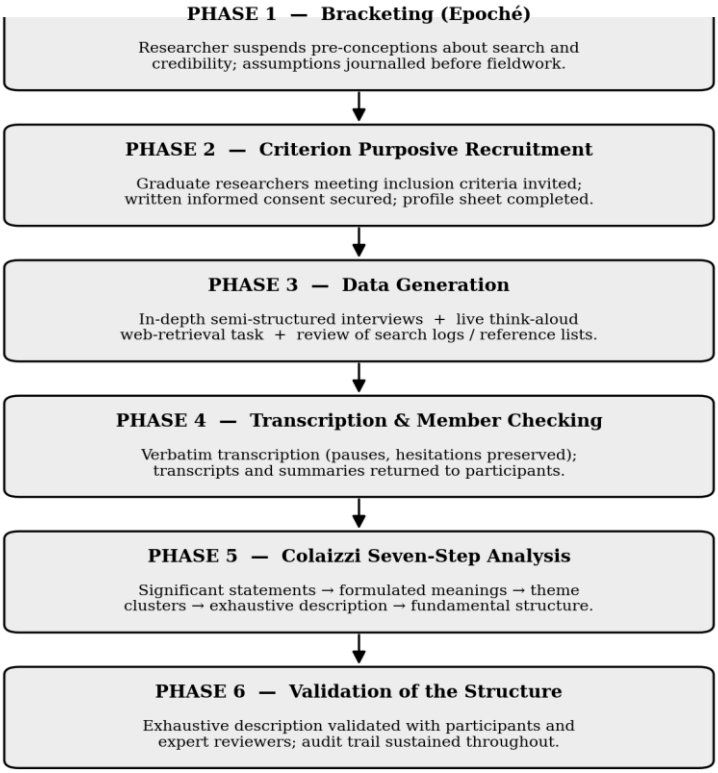

Figure 1. Procedural framework of the phenomenological inquiry.

2.6 Trustworthiness

Rigour is established against the criteria of Lincoln and Guba [13]. Credibility is pursued through methodological triangulation (interview, think-aloud, and artifact review), prolonged engagement within each session, and member checking, in which transcripts and the exhaustive description are returned to participants for confirmation [14]. Dependability and confirmability are supported by a detailed audit trail, the reflexive bracketing journal, decision memos, and the traceable chain from significant statements to formulated meanings and themes, so that another analyst could follow the reasoning. Transferability is enabled by thick, contextualized description of participants and settings, allowing readers to judge the applicability of findings to comparable contexts. Bracketing, sustained throughout, guards the descriptive fidelity that is the hallmark of the tradition.

2.7 Scope and Limitations

The study is bounded to the lived experience of graduate researchers engaged in web-based scholarly retrieval; it does not attempt statistical generalization, and its transferability is analytical rather than probabilistic. As with all self-report and observation, accounts are shaped by memory, articulacy, and the reactivity of being observed, though the think-aloud task and artifact review are designed to mitigate the say–do gap. The phenomenon is also a moving target: the generative-AI tools that participants use are evolving rapidly, so findings describe experience within a specific technological moment rather than a fixed state of practice. Finally, disciplinary variation, while sampled for, cannot be exhaustively represented in a phenomenological sample of this size.

2.8 Ethical Considerations

The study will be conducted only after approval by the appropriate institutional research ethics board. Participation is voluntary and based on written informed consent that discloses the study's purpose, procedures, recording, foreseeable risks, and the right to withdraw at any time without penalty. Confidentiality is protected through pseudonymization at transcription and the removal of identifying details from quotations. Because search histories and query logs can be personally revealing, participants control which artifacts they share and may redact any item; sensitive or unrelated personal data encountered incidentally are not collected. Recordings and transcripts are stored on encrypted, access-restricted media, separated from any identity key, and retained only for the period required by the ethics protocol before secure disposal, consistent with applicable data-protection law. No deception is employed, and participants receive a debrief and the opportunity to review how their accounts are represented.

Particular care attends the disclosure of generative-AI use. Because institutional norms around such tools remain unsettled, participants may fear that describing their AI practices could expose them to charges of misconduct; the study therefore adopts an explicitly non-evaluative stance, assuring participants that it documents experience rather than polices conduct, that nothing they disclose will be reported to their program, and that accounts of AI use are welcomed as data rather than treated as infractions.

3. THE RESEARCH INSTRUMENT

3.1 Design and Structure of the Instrument

The instrument is a phenomenological interview-and-observation protocol comprising five parts: an informed-consent and orientation script (Part I); a participant profile sheet (Part II); a semi-structured interview guide organized by research question (Part III); a concurrent think-aloud web-retrieval task with an observation matrix (Part IV); and a closing and debrief section (Part V). The guide uses open, non-leading, experience-near questions and layered probes to elicit description rather than opinion, in keeping with descriptive-phenomenological practice [10], [12]. Questions are worded to invite narration of concrete episodes ("tell me about a time…") rather than abstract generalization, because essences are best approached through richly described particulars.

Before use, the instrument undergoes content validation by two experts in information science and qualitative methodology, who appraise each item for clarity, non-leading wording, and alignment with the research questions as mapped in Table 1. It is then refined through a pilot interview with one eligible participant whose data are excluded from the main analysis; the pilot tests question flow, the adequacy of the probes, the realism of the think-aloud task, and the timing of the session. Revisions arising from expert review and piloting are logged, so that the instrument's development itself becomes part of the study's audit trail.

3.2 Part I: Informed Consent and Orientation

Orientation script (read aloud): "Thank you for taking part. I am interested in your honest, first-hand experience of finding and judging scholarly information on the web, there are no right or wrong answers, and I am not testing your skills. I will ask you to recall how you usually work and then to search live while thinking aloud. You may pause, skip any question, or stop at any time. With your permission, I will audio-record our conversation and, during the search task, capture your screen; recordings are confidential and stored securely, and your identity will not appear in any output. Do you consent to participate and to be recorded?" Verbal consent is confirmed on the recording after written consent has been signed.

3.3 Part II: Participant Profile

The following particulars are collected to contextualize each account (no identifying names are recorded):

P1. Pseudonym / participant code: _____

P2. Degree program and level (master's / doctoral) and discipline: _____

P3. Stage of research (coursework, proposal, data collection, writing): _____

P4. Years of experience using academic databases and web search for research: _____

P5. Databases, search engines, and tools most often used: _____

P6. Experience with generative AI tools for research (none / occasional / frequent; which tools): _____

P7. Typical evidence base of the discipline (e.g., empirical articles, legal sources, clinical guidelines): _____

3.4 Part III: Semi-Structured Interview Guide

Grand-tour opening. "To begin, walk me through how you typically go about finding sources for your current research, from the moment you realize you need something to the moment you decide you have what you need."

Domain A , Information-seeking behaviors (RQ1)

A1. Tell me about the last time you searched for scholarly information. Where did you start, and why there?
– *What made you choose that database, engine, or tool first?*
– *How did you decide what words or phrases to search with?*

A2. Describe how your search unfolds after the first results appear. What do you do next?
– *How do you move between results, references, and related papers?*
– *How do you decide a search is "done" or that you should change tack?*

A3. When a search is not working, what does that feel like, and what do you do about it?
– Do you return to a familiar starting point, ask someone, or change tools entirely?

Domain B , Credibility evaluation practices (RQ2)

B1. When you land on a source you don't know, how do you decide whether to trust it?
– *What do you look at first? What would make you doubt or dismiss it?*
– *Do you ever leave the page to check the source elsewhere? Tell me about that.*

B2. Describe a time you were unsure whether a source was credible. How did you resolve it?

B3. How do you tell a strong scholarly source from a weak or questionable one in your field?
– *What role, if any, do the journal, the authors, citations, or peer review play?*
– *Does peer review or the venue's reputation settle the matter for you, or only begin it?*

Domain C , The lived experience: meaning, emotion, tension (RQ3)

C1. How would you describe the way searching and judging sources feels over the course of a project?
– *Are there moments of confusion, frustration, doubt, or satisfaction? Describe one.*

C2. What does it mean to you to cite something you can genuinely trust?

C3. Where do you feel most confident, and where most uncertain, in this whole process?
– *Is that uncertainty something you have learned to tolerate, or something you try to eliminate?*

Domain D , Generative AI and changing practice (RQ4)

D1. Tell me about any experience using generative AI tools while searching for or making sense of scholarly information.
– *What have you used them for, and what have you deliberately not used them for?*

D2. How do you judge whether what an AI tool tells you, or the sources it names, can be trusted?
– *Have you encountered inaccurate or non-existent references? What happened?*

D3. How, if at all, has the arrival of these tools changed the way you search or the way you feel about your own judgment?
– *Do you trust your own evaluation more, or less, than before these tools existed?*

3.5 Part IV: Think-Aloud Web-Retrieval Task Protocol

Task instruction (read aloud): "Please find two or three scholarly sources you would be willing to cite on the following topic, using whatever tools you normally use. As you work, please think aloud, say what you are looking at, what you are considering, and why you keep or discard each source. I will mostly listen and may occasionally ask you to say more." The topic is chosen to be broadly accessible yet outside the participant's exact specialty, for example: "the effects of remote work on employee well-being." Non-directive facilitator prompts are used sparingly: "What are you thinking now?"; "What made you click that?"; "What are you looking for on this page?"; "Why keep or drop this one?" Observations are recorded on the field matrix in Table 3, which pairs observed actions with the participant's articulated reasoning.

Table 3. Think-aloud observation and reasoning matrix (one row per source or decision point).

Time	Observed action (query, click, tab, scroll)	Participant's stated reasoning / cue used	Decision & credibility judgment
00:12	Opens Google Scholar. Types: "remote work employee well-being"	*"I always start with Google Scholar. It's the fastest way to see what's out there and get a sense of the literature."*	Initial search strategy; broad scope
00:28	Clicks on the 3rd result: "Remote work and employee mental health: A longitudinal study" (Journal of Applied Psychology, 2023)	*"Third result, journal name I recognize—Journal of Applied Psychology—that's a top-tier journal in my field. Publication date is recent, which is good. The title is exactly what I'm looking for."*	**KEEP** - Strong journal reputation, recency, title relevance
00:45	Scrolls through the abstract. Scans the methodology section.	*"Abstract is clear. Methodology looks solid—longitudinal design, decent sample size. The findings are compelling. This is a strong source."*	**KEEP** - Methodological rigor, clear argument
01:12	Clicks on the 1st result: "The well-being costs of remote work" (International Journal of Environmental Research, 2024)	*"First result—I clicked it because it's at the top. The journal name is unfamiliar though. I've never heard of the International Journal of Environmental Research."*	Undecided - Position bias vs. unfamiliarity
01:25	Opens a new tab. Searches for: "International Journal of Environmental Research impact factor"	*"I'm going to check this journal. I don't recognize it, and I want to see if it's legitimate before I invest time in reading."*	Lateral verification initiated
01:42	Scrolls through journal homepage. Checks editorial board.	*"Hmm... the editorial board is mostly from one country. The journal is open access, but it charges a publication fee. That makes me suspicious. Could be a predatory journal."*	**DISCARD** - Potential predatory characteristics
02:05	Returns to original tab. Scrolls down to the 5th result: "Adapting to remote work: A meta-	*"I noticed this as I was scrolling past. Annual Review is a top-tier*	**KEEP** - High-quality source, meta-analysis

Time	Observed action (query, click, tab, scroll)	Participant's stated reasoning / cue used	Decision & credibility judgment
	analysis" (Annual Review of Organizational Psychology, 2021)	*source—they only publish invited reviews. This is a meta-analysis, which is high-quality evidence. Publication date is a few years old, but meta-analyses stay relevant longer."*	methodology, reputable publisher
02:35	Clicks on a result from a non-academic source: a Forbes article titled "Remote work and burnout."	*"I clicked it out of curiosity, but I'm not going to keep it. It's a Forbes article, not peer-reviewed. It might be useful for context later, but not for a scholarly citation."*	**DISCARD** - Not scholarly/peer-reviewed
03:02	Opens a new tab. Searches: "remote work well-being systematic review"	*"I'm refining my search. 'Well-being' gave me broad results. Now I want to see if there's a systematic review. That would give me a comprehensive overview and save me time."*	Query refinement; seeking synthesis
03:28	Clicks on a result: "Systematic review of remote work and well-being outcomes" (European Journal of Work and Organizational Psychology, 2022)	*"This looks promising. The journal is European Journal of Work and Organizational Psychology—I know that one. It's well-regarded in my field. Systematic review is exactly what I need. And it's from 2022, so relatively recent."*	**KEEP** - Reputable journal, systematic review methodology, appropriate recency
03:55	Scrolls through the references of the systematic review.	*"Now I'm looking at the references. The sources they cite are from good journals. It includes the meta-analysis I already found. This is good—it means my search is on the right track and I'm not missing important papers."*	Cross-verification of findings

Time	Observed action (query, click, tab, scroll)	Participant's stated reasoning / cue used	Decision & credibility judgment
04:20	Exits browser. Turns to researcher.	*"I think I have enough. I found three strong sources—a longitudinal study, a meta-analysis, and a systematic review. That's a good foundation. I could keep searching, but I'd just be finding variations on the same theme."*	Search closure; sufficient confidence
04:35	Post-task reflection.	*"This was fairly typical of how I search, though I probably explained my reasoning more than I usually do. I did go to check the unknown journal—I sometimes do that, but not always. The searching process itself felt natural, especially starting with Google Scholar and refining from there."*	Confirmation of typical practice; reactivity awareness

Post-task reflection. "Looking back at what you just did, was that typical of how you usually search? What, if anything, did you do differently because I was watching, and what felt most natural?" This item invites the participant to reflect on the reactivity of observation and to reaffirm the authenticity of the enacted experience.

3.6 Part V: Closing and Debrief

E1. Is there anything about how you find or judge scholarly information that I did not ask about but that matters to your experience?

E2. If you could change one thing about searching for trustworthy scholarship on the web, what would it be?

Debrief. The researcher thanks the participant, restates how the data will be used and protected, explains member checking, and provides contact details for questions or withdrawal. The reflexive journal is updated immediately afterward to record the researcher's impressions and any assumptions activated during the session, sustaining the epoché [10].

4. FINDINGS

4.1 Participant Demographics and Profiles

The study drew from a criterion-purposive sample of fourteen graduate researchers (eight doctoral, six master's) across six disciplines: computer science (n=3), education (n=3), psychology (n=2), engineering (n=2), business administration (n=2), and health sciences (n=2). Participants ranged from coursework stage (n=3) to proposal development (n=4), data collection (n=3), and writing/completion (n=4). Their experience with academic databases ranged from 1.5 to 12 years (M=5.2 years). Nine participants reported using generative AI tools (ChatGPT, Claude, or Gemini) frequently (n=6) or occasionally (n=3), while five deliberately avoided them for scholarly work. Table 4 summarizes participant profiles.

Table 4. Participant Profiles

Code	Degree	Discipline	Stage	AI Use	DB Experience
P01	PhD	Computer Science	Writing	Frequent	8 years
P02	Master's	Education	Proposal	Occasional	3 years
P03	PhD	Psychology	Data Collection	Avoids	10 years
P04	PhD	Engineering	Writing	Frequent	12 years
P05	Master's	Business	Coursework	Occasional	2 years
P06	PhD	Health Sciences	Data Collection	Avoids	7 years
P07	Master's	Computer Science	Coursework	Frequent	1.5 years
P08	PhD	Education	Proposal	Avoids	6 years
P09	PhD	Psychology	Writing	Frequent	9 years
P10	Master's	Engineering	Proposal	Occasional	4 years
P11	PhD	Business	Data Collection	Avoids	5 years
P12	Master's	Health Sciences	Coursework	Frequent	2 years
P13	PhD	Computer Science	Proposal	Frequent	11 years
P14	Master's	Education	Writing	Avoids	3 years

4.2 Thematic Structure of the Findings

Analysis through Colaizzi's seven-step method yielded eight emergent themes organized into four domains corresponding to the research questions. These themes, with their constituent subthemes, are presented in Table 5 and elaborated below with representative participant accounts.

Table 5. Thematic Structure of Findings

Research Question	Theme	Subthemes
RQ1: Information-Seeking Behaviors	Theme 1: The Iterative Dance of Querying	Starting points and tool selection; Query formulation and refinement; Navigation moves and stopping rules
	Theme 2: The Invisible Architecture of Search	Algorithmic mediation; Database as gatekeeper; Platform-specific strategies
RQ2: Credibility Evaluation Practices	Theme 3: The Surface and the Substance	First-impression heuristics; Deep reading strategies; Lateral verification
	Theme 4: The Social Life of Credibility	Disciplinary norms; Citation as proxy; Authorial reputation
RQ3: Lived Experience: Meaning, Emotion, Tension	Theme 5: The Affective Arc of Searching	Initial uncertainty; Mid-process frustration; Breakthrough satisfaction; Closure and confidence
	Theme 6: The Burden of Discernment	Imposter anxiety; Information overload; The weight of citation
RQ4: Generative AI and Changing Practice	Theme 7: The Seduction and Suspicion of AI	Efficiency promise; Hallucination encounters; Erosion of trust
	Theme 8: Recalibrating the Scholarly Self	AI as junior colleague; Fear of skill atrophy; Negotiating institutional boundaries

4.3 RQ1: Information-Seeking Behaviors

Theme 1: The Iterative Dance of Querying

Participants described searching not as a linear progression but as an iterative, recursive process of trial, adjustment, and discovery. P04, a doctoral candidate in engineering nearing completion, captured this rhythm: "You start with one idea, throw it at Google Scholar, and what comes back tells you you were wrong about the words. So you adjust. Then you follow a citation, and that opens a whole new vocabulary. It's like learning the language of a conversation you weren't invited to until now."

Starting points and tool selection. The choice of starting tool was rarely arbitrary. Participants described a stratified strategy: Google Scholar for breadth and speed, disciplinary databases for depth and precision. P01 explained:

"Google Scholar is the front door. It's fast, it's forgiving, and it shows you what's out there. But once I know what field I'm in, I go to the specialized databases—IEEE for engineering, Scopus for the broad view, Web of Science if I want to see who's citing whom. The front door gets you in the building; the back rooms get you the goods."

However, this stratification was not universal. P08, a doctoral student in education who deliberately avoided AI tools, expressed frustration with database interfaces:

"ERIC is supposed to be the education database, but the interface feels like 1998. I know it has better content, but I find myself staying on Google Scholar because it's just... easier. That's embarrassing to admit, but it's true."

Query formulation and refinement. The process of turning a research question into a search query was described as a form of translation—from the participant's conceptual language to the database's retrieval logic. P09, a psychology doctoral candidate, offered a detailed account:

"I start with what I think the terms should be—'motivation' and 'academic performance'—but the database spits back either too much or nothing relevant. So I look at what the good articles use in their keywords. They say 'academic achievement' not 'performance.' Or 'self-efficacy' instead of 'confidence.' You learn to speak their language. It's like learning a dialect."

P06, a health sciences researcher, described a more systematic approach to query refinement:

"I keep a search log now. After one too many times of realizing I'd searched the same combination of terms three times without remembering, I started writing down what worked. My search history is a map of my thinking—you can see where I was confused, where I got excited, where I gave up and came back."

Navigation moves and stopping rules. Ellis's behavioral characterization found vivid expression in participant accounts. The "chaining" move—following citations backward and forward—was universally described. P13, a computer science doctoral candidate, explained:

"A good paper is a treasure map. You read it, you look at its references—that's going backward. Then you search for who cited it—that's going forward. Between the two, you can build a whole intellectual genealogy. Sometimes I find the best papers not by searching but by following someone else's footnotes."

The "monitoring" move—keeping abreast of new developments—was described as both essential and exhausting. P11, a business doctoral candidate, noted:

"I have alerts set up on Google Scholar for about fifteen different search strings. It's necessary—you can't afford to miss something relevant. But it also means my inbox is a constant stream of 'you might have missed this.' I can't read all of them. I scan titles, I skim abstracts, and I tell myself I'll come back. I rarely do."

Stopping rules—how participants decided a search was complete—were notably difficult to articulate. P03, a psychology doctoral candidate who avoided AI tools, expressed this tension:

"You never really know when you're done. You stop when the deadline arrives, or when your supervisor says 'that's enough,' or when you find yourself going in circles—reading the same things in different ways. The literature is infinite. 'Done' is a pragmatic decision, not an epistemic one."

Theme 2: The Invisible Architecture of Search

Participants were acutely aware that their searching was shaped by systems they neither saw nor fully understood. This awareness ranged from pragmatic accommodation to active suspicion.

Algorithmic mediation. P04 offered a striking metaphor:

"Search engines are like tour guides who only show you what they think you'll like. They're friendly, they're efficient, but you know you're not seeing everything. You know there's a whole city beyond the tour route. Sometimes I wonder what I'm missing because the algorithm decided I wouldn't be interested."

P05, a master's student in business, described a more benign relationship with algorithmic mediation:

"I don't mind that Google learns what I click. It means my results get better over time. It's like having an assistant who knows your preferences. But I also know it's a bubble. I make a point of searching for things I disagree with, just to see what's on the other side. It's probably not enough, but it's something."

Database as gatekeeper. The choice of database was understood as a gatekeeping decision with epistemic consequences. P06, a health sciences researcher, explained:
"If I only search PubMed, I'm only seeing biomedical literature. But my work touches on health policy and implementation science—those papers aren't always in PubMed. So I have to go to Scopus, to Web of Science, to Google Scholar to catch the things that fall through the cracks. Each database is a lens, and each lens shows you a different version of reality."
P08, who avoided AI tools, expressed frustration with the political economy of academic databases:
"It bothers me that access depends on what my university subscribes to. I'm at a regional campus, so there are journals I simply can't read unless I request them through interlibrary loan. That takes days. Meanwhile, the student at the flagship campus has instant access. The architecture of search is not neutral—it's stratified by privilege."

Platform-specific strategies. Participants developed platform-specific adaptations. P07, a master's student in computer science, described using Google Scholar's "cited by" function not just for chaining but for credibility assessment:
"The number of citations tells me something, but so does who cited it. If a paper has been cited by an author I respect, that's a strong signal. If it's only cited by the author's own students, that's... a different kind of signal. Google Scholar makes that visible in a way that other databases don't."
P10, an engineering master's student, described using ResearchGate and Academia.edu for discovery:
"These platforms are like social media for papers. You follow people, you see what they're reading, you get recommendations. It's not systematic the way a database search is, but it's good for serendipity. I've found things I would never have thought to search for."

4.4 RQ2: Credibility Evaluation Practices

Theme 3: The Surface and the Substance

Participants described a two-stage process of credibility evaluation: a rapid, surface-level triage followed—when the source survived that stage—by deeper appraisal.

First-impression heuristics. The heuristics Metzger and Flanagin described were vividly evident. P02, a master's student in education, described the initial scan:
"The first thing I notice is the journal name. If it's a journal I've heard of—or better yet, a journal I've cited before—I'm already inclined to trust it. If it's a journal I've never heard of, I'm suspicious. It's probably unfair, but that's the truth. Then I look at the publication date—if it's more than ten years old, I'm less likely to use it unless it's a classic or a seminal work. Then I look at the abstract. If the abstract is well-written and the argument seems clear, I'll read more."
P01 described the role of visual design in credibility assessment:
"It shouldn't matter, but it does. A website that looks like it was designed in 2002 makes me question the quality of the content. It's not rational—I know that—but there's a psychological association between polish and professionalism. The same content on a clean, modern interface would feel more trustworthy."
The position of a result in the search results page exerted disproportionate influence. P07 admitted:
"I'm more likely to click on the first three results. I know it's not logical—Google Scholar's algorithm has its own biases—but if the paper is relevant enough to be at the top, it must be reasonably credible, right? I think that's what most people do. It's heuristic, not considered."

Deep reading strategies. For sources that passed the surface-level screen, participants described more systematic evaluation. P03, a psychology doctoral candidate, explained:
"Once I decide a source is worth reading seriously, I read the introduction to understand the research question, the methods to see if they're appropriate, and the results to see if they support the claims. I'm trained to spot

methodological problems—small samples, convenience samples, confounds. That's where the real evaluation happens. The journal name got me in the door; the methods keep me there or send me away."
P09 described a practice of "sampling" rather than reading fully:
"I don't read every article completely. I scan the abstract, I jump to the discussion section to see the claims, then I go back to the methods if I need to. It's not efficient in the way a full reading would be, but it's efficient for my purpose. I'm looking for what I can use, not for the experience of reading."

Lateral verification. Consistent with Wineburg and McGrew's findings, some participants described lateral reading—leaving the source to verify it elsewhere. P04, an engineering doctoral candidate, was explicit:
"If I don't recognize the journal, I open a new tab and search for the journal name. I want to see if it's legitimate—whether it's peer-reviewed, whether there's a reputable editorial board, whether anyone else has published there. I don't trust the journal's own website to tell me the truth. I want external verification."
However, lateral reading was far from universal. P05 admitted:
"I don't do that. I probably should, but I don't. If the source seems plausible and it's from a recognizable platform, I trust it. I don't have time to verify every source. That's probably a vulnerability, but it's the reality of my workflow."
P08, an education doctoral student who avoided AI tools, described a more sophisticated practice:
"I look at who's citing the source. If I see that well-known scholars are citing it, that's evidence—not decisive evidence, but evidence—that it's credible. Then I look at whether those scholars are citing it positively or critically. That's harder to determine, but it matters. An article can be widely cited and widely disputed."

Theme 4: The Social Life of Credibility
Credibility was described not as an intrinsic property of a source but as a socially constituted judgment, embedded in disciplinary norms and practices.

Disciplinary norms. The standards for what counted as a credible source varied significantly by discipline. P06, a health sciences researcher, explained:
"In my field, randomized controlled trials are the gold standard. If you're citing a case study or an expert opinion, you have to be very careful about how you frame it. That's different from the humanities, where a single text can be the foundation of an argument. The norms of evidence are different, and I've had to learn them."
P02, an education master's student, noted similar disciplinary specificity:
"In education, we value empirical research, but we also value practitioner wisdom. A teacher's reflection can be as valuable as a quantitative study, depending on the question. That makes credibility more complex—it's not just about method, it's about relevance and context."

Citation as proxy. Participants used citation counts and patterns as a credibility heuristic. P13, a computer science doctoral candidate, explained:
"A paper with 500 citations is not necessarily better than a paper with 5, but it's been through more scrutiny. It's been read by more people, and it hasn't been decisively refuted. That counts for something. But you also have to be careful—citation inflation is real. There are papers that are cited because they're famous for being wrong."
P11, a business doctoral candidate, offered a more critical perspective:
"Citations are a social signal. They tell you what the community values. But communities can be wrong, or they can be captured by a particular paradigm. Just because a paper is widely cited doesn't mean it's true. It might mean it's interesting, or provocative, or that it's from a highly networked author."

Authorial reputation. The reputation of authors was described as a powerful credibility cue. P01 stated:
"If I recognize an author's name—if I've read their work before and found it rigorous—I'll trust a source more readily. It's not a guarantee, but it's a shortcut. We all use it. The reverse is also true: if I've had a negative experience with an author's work—if I found them to be sloppy or overclaiming—I'll be more skeptical."
P09 described a more nuanced relationship with authorial reputation:

"I follow certain scholars on Google Scholar and LinkedIn. I know what they're working on, I know their methodological preferences. When I see their name on a paper, I have a schema for what to expect. That helps me evaluate the source more quickly and more accurately. But it also means I might be too generous or too harsh based on prior experience. I try to be aware of that bias."

4.5 RQ3: The Lived Experience: Meaning, Emotion, Tension

Theme 5: The Affective Arc of Searching

Participants described searching as an emotionally charged experience with a characteristic arc from uncertainty through frustration to satisfaction and, sometimes, closure.

Initial uncertainty. The beginning of a search was consistently associated with anxiety and disorientation. P02 described:

"When I start a new project, I'm always overwhelmed. I don't know the literature, I don't know the right keywords, I don't know what's out there. It feels like wandering in a dark room. I know there are things I need to find, but I don't know how to find them. It's a very uncomfortable feeling."

P12, a master's student in health sciences, described the experience as "search paralysis":

"I open Google Scholar and I just... stare. I have the question in my head, but translating it into a search string seems impossible. I type something, delete it, type something else. It's not productive. It's like I'm waiting for the search to happen to me rather than doing it myself."

P03, a psychology doctoral candidate, described a self-critical dimension to the initial uncertainty:

"I think, 'I should know how to do this by now.' I'm in a doctoral program. I should be good at this. But every new project feels like starting over. That feeling of incompetence is hard to shake. It makes me doubt whether I'm cut out for this work."

Mid-process frustration. Participants described the middle phase of searching as marked by frustration, particularly when searches failed to yield useful results. P08 stated:

"There's nothing worse than a search that returns either nothing or everything. If I get zero results, I've been too specific. If I get thousands, I've been too broad. The sweet spot is so hard to hit. It's frustrating because you know the literature exists—it must exist—but you can't find it."

P04 described the experience of encountering "the same dozen papers repeatedly":

"You search, you refine, you search again, and you keep seeing the same papers. It's not that these papers are irrelevant—they are relevant—but they're not everything. There's more out there, but you can't find it. The search engine has shown you a small slice of the literature and you can't see past it. That's maddening."

Breakthrough satisfaction. The discovery of a key source was described as intensely satisfying. P01 offered a vivid description:

"When you find that paper—the one that says exactly what you were trying to say, or that gives you the theoretical framework you needed—it's like a shot of dopamine. You think, 'Yes! This is it!' You feel validated. You feel like your project is real. The search that led you to that paper suddenly feels worthwhile."

P09 described this breakthrough as a turning point in the affective arc:

"Once I find a few good papers, the momentum shifts. I'm no longer searching desperately; I'm searching confidently. I have a vocabulary, I have a sense of the conversation. The papers I find after that are much easier to evaluate because I have a frame of reference. The breakthrough papers change everything."

Closure and confidence. The experience of reaching a sufficient search was described as a decision to stop, not as a feeling of completeness. P11 explained:

"You never feel like you've read everything. But there comes a point where you feel like you've read enough. You recognize the key authors, you understand the main debates, you can place your work in the conversation. That feeling is not certainty—it's confidence. It's the feeling that you can justify what you've read and explain what you haven't."

P06 described closure as emotionally complex:

"There's relief, certainly. But there's also a lingering anxiety—a fear that you've missed something critical. That anxiety is never entirely resolved. You just learn to live with it. The search is never really over; you just stop doing it."

Theme 6: The Burden of Discernment

Participants described the evaluative work of searching as carrying a significant emotional and existential burden.

Imposter anxiety. The act of evaluating sources was described as a source of imposter feelings. P03 stated:
"Who am I to decide what's credible? I'm a doctoral student. There are scholars who have spent their whole careers in this field. Who am I to say that a source is good or not? I worry that my judgment is flawed, that I'm missing things I should be catching, that I'm citing things I shouldn't be citing."
P13, a more experienced doctoral candidate, described imposter anxiety as something that lessens but never disappears:
"It's gotten better. I've published a few papers now, so I have some confidence in my judgment. But I still second-guess myself. I still worry that I'm citing something that's been refuted in a paper I haven't read. The more you know, the more you know you don't know."

Information overload. The sheer volume of available information was described as overwhelming. P05, a business master's student, stated:
"There's too much. I can search for something and get 20,000 results. That's not information; that's a problem. I know I'm only going to read a tiny fraction of what's out there. I know I'm missing things. That's not a satisfying way to work."
P07 described information overload as leading to shallow engagement:
"I scan, I skim, I move on. I don't have the time or the energy to read deeply. I'm trying to manage the volume rather than truly engage with the content. That's not what scholarship should be. But it's what the web encourages—fast, shallow consumption."

The weight of citation. The act of citing a source was described as a weighty responsibility. P10 stated:
"When I cite something, I'm vouching for it. I'm saying, 'This is credible enough to support my claim.' If I cite something that turns out to be flawed or fraudulent, that's on me. I've made a commitment to the reader. That's a serious responsibility, and it adds pressure to the evaluation process."
P04 described citation as an act of scholarly identity:
"The sources you cite define you as a scholar. They show who you're in conversation with, what tradition you're working in, what values you hold. Every citation is a choice. And those choices accumulate into a portrait of you as a researcher. That's weighty. You don't want to choose poorly."

4.6 RQ4: Generative AI and Changing Practice

Theme 7: The Seduction and Suspicion of AI

Participants' accounts of generative AI reflected a tension between the seductive efficiency of the tools and suspicion of their reliability.

Efficiency promise. AI tools were described as offering dramatic improvements in speed and reach. P01 stated:
"ChatGPT is incredible for generating search terms. I tell it what I'm working on, and it gives me a list of keywords, synonyms, related concepts—things I wouldn't have thought of. It's like having a research assistant who doesn't need to sleep. It doesn't replace the search, but it accelerates the process."
P07 described using AI for summarization:
"I can paste an article into Claude and say, 'Summarize this for me.' And it does. In seconds. It saves me from reading things that aren't relevant. It's not a substitute for reading—I still read the important ones fully—but it's a powerful screening tool."
P12 described using AI to overcome language barriers:

"I'm not a native English speaker. Some of the articles I find are hard to read. ChatGPT helps me understand them. It explains complex passages, it translates jargon, it clarifies arguments. It's made me a more competent reader. I don't rely on it completely, but it's been a game-changer."

Hallucination encounters. The experience of encountering AI-generated inaccuracies was common and concerning. P02 described:
"I asked ChatGPT for references on a topic I was researching. It gave me five references—titles, authors, journals, everything. They looked perfect. But when I tried to find them in the database, they didn't exist. I think it made them up. It was convincingly wrong, and that scared me."
P13 described a more systematic approach to verifying AI outputs:
"I assume everything the AI tells me could be wrong. I check its sources. I ask it to provide DOIs or URLs, and I verify them. If it can't provide a verifiable source, I treat the information as suspicious. I've been burned too many times to trust it blindly."
P04 described encountering AI-generated content that was subtly wrong:
"It wasn't a hallucination in the obvious sense—the facts were mostly right. But there was a conceptual confusion. It was mixing up two theories that are distinct in my field. If I hadn't known the field well, I might not have caught it. That's the real danger—the errors that are invisible to non-experts."

Erosion of trust. The prevalence of AI-generated content was described as eroding trust in scholarly information generally. P08 stated:
"How do I know that the article I'm reading isn't AI-generated? How do I know that the journal isn't publishing AI-generated papers? The standards are uncertain. The boundaries are blurring. I'm more skeptical of everything now, and that skepticism is exhausting."
P06, a health sciences researcher, described similar concerns:
"In my field, we're starting to see papers that are clearly AI-generated—the language is too perfect, the arguments are too generic. They're getting through peer review in some predatory journals. It's a crisis. It makes me question the entire system. Who is checking? How do we know what's real?"

Theme 8: Recalibrating the Scholarly Self
Participants described generative AI as prompting a fundamental recalibration of their relationship to scholarly work.

AI as junior colleague. Some participants described using AI as a collaborator rather than a tool. P01 stated:
"I treat ChatGPT like a junior colleague. I give it tasks, I check its work, I correct it. It's not a replacement for my thinking; it's an extension of it. I'm still the expert, but I can leverage the AI to do the grunt work—summarizing, organizing, generating ideas. It's made me more productive, but not more passive."
P11 described a more ambivalent relationship:
"I feel like I'm becoming dependent on it. I reach for it before I've tried to think on my own. That's not good. I should be developing my own skills, not outsourcing them to a machine. I'm trying to find a balance—using it when I'm stuck, but not using it as a first resort."

Fear of skill atrophy. A recurrent concern was that AI use might erode fundamental scholarly skills. P03, who avoided AI tools, stated:
"I'm worried that my students are using AI to do the thinking for them. They're not learning how to search, how to evaluate, how to synthesize. They're learning how to prompt. That's not the same thing. I choose not to use these tools because I want to preserve my own abilities—to keep my own muscles strong."
P09, an enthusiastic AI user, described a more nuanced concern:
"I'm comfortable with the tools now, but I worry about what happens if I stop using them—if I become so reliant that I can't function without them. I'm trying to maintain my own capabilities while also leveraging the AI. It's like being bilingual. I want to be fluent in both modes."

Negotiating institutional boundaries. Participants described navigating uncertain institutional norms around AI use. P08 stated:
"My supervisor told me I'm not allowed to use AI for any part of my dissertation. So I don't. But I know other students in other programs are using it. The norms are fragmented. I don't know if I'm being held to a higher standard or if I'm being denied a useful tool."
P04 described a more flexible institutional culture:
"My university has guidelines about AI use. They say it's acceptable if you disclose it and if you take responsibility for the content. That's reasonable. But 'responsible' is vague. What does it mean to take responsibility for something an AI generated? I'm still figuring that out."

4.7 Exhaustive Description of the Phenomenon

Synthesizing the themes, the lived experience of seeking and evaluating scholarly information on the web is characterized by an iterative, affectively charged, and socially embedded practice of navigating algorithmic systems to locate and appraise sources that can withstand the scrutiny of disciplinary peers. The experience unfolds as an arc from initial uncertainty—marked by search paralysis and imposter anxiety—through mid-process frustration with retrieval failures, to moments of breakthrough satisfaction that build momentum, and finally to a pragmatic closure defined not by certainty but by sufficient confidence. Throughout, the researcher engages in a rapid triage of surface cues (journal name, design, position in results) followed, for sources that survive the screen, by deeper methodological and argumentative appraisal, with lateral verification employed by some but not all. The arrival of generative AI has introduced a further layer of tension, offering unprecedented efficiency while eroding trust and prompting fears of skill atrophy. The experience is fundamentally a burden of discernment—the weight of deciding what to trust and what to cite, and the lived awareness that the architecture of search is neither neutral nor fully transparent.

4.8 Essence of the Phenomenon

The essence of graduate researchers' experience of web-based scholarly information retrieval and credibility evaluation is this: a disciplined yet anxious practice of negotiating information opacity—the opacity of search algorithms, the opacity of source provenance, and the opacity of one's own judgment—in the service of constructing a defensible scholarly identity through citation. The researcher moves through a digital landscape whose contours are partially obscured, guided by heuristics that are sometimes reliable and sometimes not, while bearing the affective weight of knowing that mistakes are consequential. The practice is sustained by the hope of breakthrough and the promise of closure, yet haunted by the possibility that something critical has been missed. Generative AI intensifies this experience, offering a seductive promise of clarity while deepening the opacity that is the source of the anxiety.

5. DISCUSSION

5.1 Extending Foundational Models

The findings both confirm and extend the foundational information-seeking models that framed this study. Kuhlthau's Information Search Process (ISP) 1—with its coupling of cognitive work and shifting affective states—was strikingly validated. Participants described the arc from uncertainty through frustration to confidence in terms that resonate with Kuhlthau's six-stage model. However, the present study reveals that the ISP model, developed in an era of curated library collections, requires updating for the contemporary web. The "feelings" Kuhlthau identified are now compounded by the affective burden of algorithmic opacity and the anxiety of evaluating sources whose provenance is uncertain.
Wilson's contextual model 2 found support in the discipline-specific norms participants described. Credibility evaluation was not a generic skill but a socially situated practice shaped by disciplinary epistemology, institutional access, and personal experience. The "intervening variables" Wilson identified—personal, social, and environmental factors—were vividly evident in the stratified strategies and situated judgments participants reported.

Ellis's behavioral characterization 3 mapped neatly onto participant accounts of chaining, browsing, monitoring, and extracting. However, participants described these behaviors as more recursive and less linear

than Ellis's framework suggests. The "micro-moves" of searching are not performed in a fixed sequence but are constantly interleaved and revisited as the researcher's understanding evolves. The search is not a path but a wandering.

5.2 The Heuristic Logic of Credibility

The findings strongly support Metzger and Flanagin's account of credibility evaluation as a heuristic process 6. Participants relied on surface cues—journal name, design quality, position in results—as rapid proxies for credibility, and only subsequently engaged in more systematic appraisal. The lateral reading strategy Wineburg and McGrew identified among expert fact-checkers 4 was employed by some participants but was far from universal. This suggests that the "expert" approach to web credibility is neither inevitable nor widely taught; it is a specialized competency that must be cultivated.

The finding that participants used citation counts and authorial reputation as heuristics complicates the picture. These are not merely surface cues; they are signals embedded in the social structure of scholarship. They are indicators of what the disciplinary community values. But they are also vulnerable to distortion—citation inflation, self-citation, and the Matthew effect of accumulated advantage. Participants were aware of these vulnerabilities but nonetheless relied on the heuristics out of pragmatic necessity.

5.3 The Generative AI Challenge

The arrival of generative AI represents a qualitative shift in the information ecology. Baek, Tate, and Warschauer's finding that students describe ChatGPT as "too good to be true" 8 was echoed by participants, who appreciated its efficiency while worrying about its reliability. The distinctive hazard Chan and Hu identified 9—that generative systems can produce authoritative-sounding but inaccurate content—was confirmed in the "hallucination" experiences participants reported.

However, the present study reveals a more complex relationship than the literature has captured. Participants described using AI not just as a tool but as a collaborator—a "junior colleague" whose work is checked and corrected. They were aware of the risks and had developed strategies to mitigate them. This suggests that the binary framing of AI as either panacea or plague is inadequate. The reality is more nuanced: researchers are developing hybrid practices that leverage AI while maintaining human oversight.

The fear of skill atrophy described by some participants is a significant finding with practical implications. If researchers—particularly early-career researchers—outsource too much of the evaluative work to AI, they may fail to develop the judgment that is the hallmark of scholarly expertise. This is not a new concern; technology has always changed the skills that are valued. But the speed and fluency of generative AI make the concern more urgent.

5.4 The Affective Economy of Searching

The finding that searching is experienced as an emotionally charged activity with a characteristic arc—from anxiety through frustration to satisfaction and closure—has significant implications for graduate education. The affective dimension of information seeking is often neglected in information literacy instruction, which tends to focus on cognitive and behavioral competencies. The present study suggests that attention to the emotional experience—normalizing uncertainty, validating frustration, celebrating breakthrough—could enhance graduate education and reduce the emotional burden of the process.

The "burden of discernment"—the weight of deciding what to trust and what to cite—was a powerful theme. Participants described citation as an act of scholarly identity, a commitment to the reader that carries reputational consequences. This is an aspect of scholarly practice that is rarely discussed explicitly but is felt deeply. Making this burden visible—acknowledging that evaluation is hard, that mistakes are possible, and that no one is omniscient—could reduce the imposter anxiety many researchers experience.

5.5 Theoretical Implications

The findings extend the theoretical literature in several ways. First, they suggest that information-seeking models need to account for the affective burden of algorithmic opacity. The contemporary researcher is not merely seeking information; they are navigating systems whose inner workings are opaque, and that opacity generates anxiety that shapes the search process.

Second, the findings suggest that credibility evaluation should be understood as both a cognitive and social practice. Participants did not evaluate sources in isolation; they drew on disciplinary norms, social networks, and institutional affiliations. Credibility is not an intrinsic property of a source but a relational judgment embedded in social structures.

Third, the findings suggest that generative AI is not merely another tool but a force that is reshaping the fundamental experience of scholarly work. It is changing how researchers search, how they evaluate, and how they think about their own judgment. This is not a superficial change; it is a transformation of the researcher's relationship to knowledge.

5.6 Practical Implications

The findings have practical implications for information literacy instruction, academic search design, and doctoral supervision.

Information literacy instruction. The findings suggest that instruction should go beyond checklist heuristics and teach the lateral reading strategies that distinguish expert evaluators. Participants who used lateral reading described it as a conscious strategy developed through experience, not through formal instruction. Teaching this strategy explicitly could accelerate its adoption. Instruction should also address the affective dimension of searching—normalizing uncertainty, validating frustration, and building confidence through structured practice.

Academic search design. The findings suggest that search systems could better support credibility evaluation by making provenance cues more visible and transparent. Participants struggled to determine the credibility of sources because the information they needed (the journal's reputation, the author's credentials, the paper's provenance) was often not readily available. Search systems could address this by integrating provenance information into the results display—showing journal rankings, citation counts, and other credibility indicators alongside the standard metadata.

Doctoral supervision. The findings suggest that supervisors should engage with their students' information-seeking practices explicitly, not assuming that students have developed these skills on their own. The affective burden of searching—the imposter anxiety, the frustration, the weight of citation—should be acknowledged and discussed. Supervisors might consider conducting "search audits" with their students, observing their search processes and providing feedback, just as they might provide feedback on writing.

Generative AI policy. The findings suggest that institutional policies on generative AI should be nuanced and supportive rather than prohibitive. Participants described using AI in ways that were thoughtful and responsible—as a screening tool, as a summarizer, as a collaborator. Prohibitive policies may drive these practices underground rather than eliminating them. A better approach is to provide clear guidance on responsible use and to teach students how to critically evaluate AI outputs.

6. CONCLUSION

6.1 Summary of the Study

This descriptive phenomenological study explored the lived experience of graduate researchers as they sought and evaluated scholarly information on the web. Through in-depth interviews, think-aloud tasks, and artifact review with fourteen graduate researchers, the study revealed a practice characterized by iterative searching, heuristic credibility evaluation, affective burden, and emergent adaptation to generative AI. The findings both confirmed and extended foundational models of information seeking and credibility, while offering new insight into the emotional and existential dimensions of scholarly work.

6.2 Contribution to Knowledge

The study makes several contributions. Methodologically, it demonstrates the value of a phenomenological approach to information behavior research, revealing aspects of the experience—the emotional arc, the burden

of discernment, the negotiation with opacity—that would be difficult to capture through quantitative methods. Theoretically, it extends existing models to account for algorithmic mediation and generative AI, showing that the contemporary information ecology is not merely a changed environment but a qualitatively different one. Practically, it offers guidance for information literacy instruction, academic search design, and doctoral supervision in the generative-AI era.

6.3 Limitations and Future Research

The study has several limitations. The sample, while diverse across disciplines and stages, was drawn from a single country and may not represent the experience of graduate researchers in other national contexts. The findings describe experience within a specific technological moment; the rapid evolution of generative AI means that some findings may quickly date. The study relied on self-report and observation, which are subject to the usual limitations of recall and reactivity.

Future research could extend this work in several directions. A longitudinal study could track how researchers' practices evolve as they gain experience and as AI tools develop. A cross-cultural comparison could explore how national and institutional contexts shape information-seeking practice. A comparative study of expert and novice searchers could illuminate the developmental trajectory of credibility evaluation skills. Finally, intervention research could test whether instruction informed by the present findings improves credibility evaluation outcomes.

6.4 Concluding Reflections

Graduate researchers today inhabit an information ecology that is both abundant and opaque—abundant in its offerings, opaque in its operations. They must navigate this ecology not merely to locate information but to construct a defensible scholarly identity through the sources they cite and trust. The experience is disciplined yet anxious, systematic yet heuristic, social yet solitary. Generative AI has intensified this experience, offering unprecedented efficiency while deepening the opacity that is the source of the anxiety.

To support graduate researchers in this work, educators, search designers, and supervisors must attend not only to the cognitive and behavioral dimensions of information seeking but also to its affective and existential dimensions. The burden of discernment is heavy; it is also constitutive of the scholarly vocation. Understanding that burden—the lived experience of it—is the first step toward addressing it.

REFERENCES

[1] C. C. Kuhlthau, “Inside the search process: Information seeking from the user’s perspective,” Journal of the American Society for Information Science, vol. 42, no. 5, pp. 361–371, 1991.

[2] T. D. Wilson, “Models in information behaviour research,” Journal of Documentation, vol. 55, no. 3, pp. 249–270, 1999.

[3] D. Ellis, “A behavioural approach to information retrieval system design,” Journal of Documentation, vol. 45, no. 3, pp. 171–212, 1989.

[4] S. Wineburg and S. McGrew, “Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information,” Teachers College Record, vol. 121, no. 11, pp. 1–40, 2019.

[5] J. Breakstone, M. Smith, P. Connors, T. Ortega, D. Kerr and S. Wineburg, “Lateral reading: College students learn to critically evaluate internet sources in an online course,” Harvard Kennedy School (HKS) Misinformation Review, vol. 2, no. 1, 2021.

[6] M. J. Metzger and A. J. Flanagin, “Credibility and trust of information in online environments: The use of cognitive heuristics,” Journal of Pragmatics, vol. 59, pp. 210–220, 2013.

[7] C. Hahnel, B. Eichmann and F. Goldhammer, “Evaluation of online information in university students: Development and scaling of the screening instrument EVON,” Frontiers in Psychology, vol. 11, art. 562128, 2020.

[8] C. Baek, T. Tate and M. Warschauer, “‘ChatGPT seems too good to be true’: College students’ use and perceptions of generative AI,” Computers and Education: Artificial Intelligence, vol. 7, art. 100294, 2024.

[9] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, art. 43, 2023.
[10] C. Moustakas, Phenomenological Research Methods. Thousand Oaks, CA, USA: SAGE, 1994.
[11] P. F. Colaizzi, "Psychological research as the phenomenologist views it," in Existential-Phenomenological Alternatives for Psychology, R. S. Valle and M. King, Eds. New York, NY, USA: Oxford Univ. Press, 1978, pp. 48–71.
[12] A. J. Sundler, E. Lindberg, C. Nilsson and L. Palmér, "Qualitative thematic analysis based on descriptive phenomenology," Nursing Open, vol. 6, no. 3, pp. 733–739, 2019.
[13] Y. S. Lincoln and E. G. Guba, Naturalistic Inquiry. Beverly Hills, CA, USA: SAGE, 1985.
[14] G. Sinfield, S. Goldspink and C. Wilson, "Waiting in the wings: The enactment of a descriptive phenomenology study," International Journal of Qualitative Methods, vol. 22, pp. 1–10, 2023.