Open Scholarship of Teaching and Learning
osotl.org
ISSN 2752-4116

2026, Vol. 5(2) 26-54

Large Language Model–assisted curriculum audit in psychology: A prompt-based method

Angelos Rodafinos[1]

Abstract

Curricula in psychology and other disciplines face increasing pressure to align intended learning outcomes with assessment evidence while responding to changing professional demands and the growing presence of generative artificial intelligence (GenAI) in higher education. This methodological paper introduces the Prompt Playbook, a prompt-supported workflow for programme-level curriculum audit and competency mapping, with an optional later stage of the process for redesign, designed to structure faculty deliberation and improve documentation for quality assurance. The Playbook specifies three steps: (1) Programme Audit to surface misalignments among programme aims, course learning outcomes, and assessments; (2) Illustrative Competency Mapping to prioritise knowledge, skills, and abilities using external signals triangulated with disciplinary standards; and (3) Curriculum Integration to revise learning outcomes and design aligned learning activities and assessments, including paired AI-supported and AI-restricted tasks. A governance layer addressing ethics, data protection, equity, and validation is embedded through logging and verification protocols. Examples in the main text are primarily illustrative and are intended to demonstrate the structure of outputs rather than report new empirical findings; in addition, a supplementary worked example presents one actual captured LLM run to show how the validation protocol handles imperfect responses. The paper concludes with implementation pathways, limitations, and a pilot-study design for future evaluation.

Keywords

curriculum audit, constructive alignment, generative AI, assessment design, psychology education

[1] Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:

Angelos Rodafinos, School of Physical Education & Sport Science, Aristotle University of Thessaloniki, Thessaloniki, 57001 Thermi, Greece

Email: ARodafinos@gmail.com

Introduction

In psychology and related fields, graduates are increasingly expected to demonstrate digital and data literacy, ethical judgement, and competent human–technology interaction. Yet programme structures and assessment practices often lag shifts in professional practice, technological change, and ethical imperatives, creating misalignment between intended outcomes and assessment evidence (Kulasegaram et al., 2018; Mah & Groß, 2024).

In parallel, GenAI systems based on large language models (LLMs) are rapidly entering teaching, learning, and assessment workflows. While these systems can support drafting, feedback, and curriculum design activities, they also raise well-documented concerns regarding validity, bias, transparency, and the reliability of generated content (Bender et al., 2021; Dratsch et al., 2023; UNESCO, 2023). These developments intensify the need for curriculum review processes that are both methodologically structured and ethically governed.

In this context, the Prompt Playbook uses structured prompting as a form of disciplined inquiry: carefully designed prompts function as questions that help curriculum teams surface assumptions, diagnose misalignments, and identify gaps. The Prompt Playbook offers a prompt-supported workflow that deliberately separates curriculum diagnosis from curriculum design. Step 1 supports programme audit, Step 2 supports illustrative competency mapping and prioritisation, and Step 3 provides an optional, follow-up design pathway for translating validated findings into curriculum revisions, including AI-aware assessment design. In this sense, the Playbook does not treat audit and redesign as interchangeable activities: audit and mapping generate structured decision-support artifacts for faculty deliberation, while redesign occurs only as a subsequent, human-governed response to those findings. The paper presents the Playbook as a methodological contribution, with reusable prompts, documentation templates, and an auditable governance and validation protocol to support disciplined faculty deliberation. Where generated outputs are contested, the framework assumes that disagreement is resolved through ordinary academic governance rather than through deference to the tool: outputs are checked against programme evidence, disciplinary standards, and faculty judgement before any recommendation is adopted.

The paper also includes a supplementary worked example based on one actual captured LLM output, annotated to show how the validation protocol handles imperfect responses. Because large language model outputs may vary across runs, sessions, tool providers, and model updates, the protocol does not assume identical textual reproduction. Instead, it is designed to support procedural reproducibility (the same documented inputs, prompts, and checks), auditability (a traceable record of what was generated and why), and decision robustness (whether curriculum judgements remain stable despite limited variation in outputs).

Although the framework is designed for cross-disciplinary adaptation, the paper foregrounds psychology because the discipline is simultaneously navigating methodological modernisation (e.g., data skills and research practices), professional competencies (e.g., ethical judgement and reflective practice), and the pedagogical implications of GenAI. The examples and prompt artifacts are therefore oriented to common psychology curriculum contexts while remaining transferable to other fields.

The Prompt Playbook emerged from the author's reflections on curriculum review activities within education during a period of rapid growth in GenAI technologies. While numerous discussions highlighted the importance of AI literacy, practical guidance for translating these discussions into curriculum change remained limited. The framework was therefore developed as a structured mechanism for helping programme leaders move from broad aspirations regarding AI readiness to concrete curriculum decisions that could be implemented within existing Quality Assurance (QA) processes.

Learning-Theory Rationale for the Prompt Playbook

Although the Prompt Playbook includes a later stage for curriculum redesign, it is primarily presented as a curriculum-review and decision-support method. It is informed by psychological research on learning and self-regulation, particularly the idea that explicit goals, criteria, feedback loops, and structured reflection support effective judgement and action (Butler & Winne, 1995; Pintrich, 2000; Zimmerman, 1990). In this sense, the Playbook treats curriculum review as guided professional learning, using prompts as scaffolds that help faculty articulate rationales, check evidence, and coordinate decisions (van de Pol et al., 2010; Wood et al., 1976).

The remainder of the paper is structured as follows. The next subsection clarifies the paper's Contributions and Scope. Subsequent sections describe the development of the Prompt Playbook and then detail the three steps of the method (Programme Audit, Illustrative Competency Mapping, and Curriculum Integration). The paper then addresses governance requirements – ethics, data governance, and equity – followed by a section that specifies a reproducibility and validation protocol for logging, verification, and auditability. The final sections discuss implementation considerations, limitations, directions for future research, and concluding reflections.

Contributions and Scope

In terms of contribution type, this article is a methodological paper. It proposes a replicable workflow for programme-level curriculum audit and competency mapping in psychology, with an optional subsequent design phase for curriculum redesign. The primary contribution is therefore methodological and diagnostic: the Playbook is designed to help curriculum teams review the curriculum as it currently exists, identify alignment gaps and priority competencies, and document their reasoning in auditable form before any redesign decisions are made.

The aim is not to provide a comprehensive synthesis of empirical findings on curriculum reform or GenAI; rather, it is to translate established curriculum-design principles into an actionable procedure that curriculum teams can pilot, adapt, and evaluate within existing QA cycles.

The paper makes four contributions. First, it specifies a three-step method that distinguishes analytic review from design response: Step 1 (Programme Audit) establishes an evidence-informed baseline of the curriculum as currently enacted; Step 2 (Illustrative Competency Mapping) identifies and prioritises candidate Knowledge, Skills and Abilities (KSAs) through transparent triangulation; and Step 3 (Curriculum Integration with AI-aware assessment design) translates selected priorities into revised outcomes, activities, and assessments only after review-stage outputs have been validated and discussed. Second, it provides a structured set of prompts (Prompts A–L) intended to standardise how curriculum teams elicit summaries, surface alignment gaps, generate draft

refinements, and document rationales, while preserving human judgement as the locus of decision-making. Third, it introduces an explicit governance layer (Ethics and Data Governance; Implementation Considerations) and a reproducibility/validation protocol to address known risks of large language model use – such as plausible but incorrect outputs, bias, drift, and output instability across runs or models. The protocol is not limited to logging and verification; it also clarifies how teams should interpret variability, when repeat-run or cross-model checks are warranted, and how to judge whether conclusions are robust enough for curriculum deliberation. In this framework, reproducibility is defined less as exact textual sameness and more as the ability of another team to reconstruct the procedure, inspect the evidence trail, and determine whether the same broad curriculum judgement would be reached under comparable conditions. Fourth, it outlines an evaluation pathway, including a pilot study design, to enable future empirical testing of feasibility, mapping stability, workload implications, and educational value.

Throughout, the Prompt Playbook outputs are framed as decision-support artifacts – inputs to structured faculty deliberation – rather than prescriptions. This applies especially to Steps 1–2, whose purpose is to support QA-oriented review, alignment checking, and prioritisation. Step 3 is not presented as audit proper, but as a possible design-oriented extension that departments may pursue once review findings have been interpreted through human judgement and programme governance.

Development of the Prompt Playbook

The Prompt Playbook was developed iteratively as a structured review tool for programme- and course-level curriculum work in psychology, with a subsequent design component. It combines curriculum-review principles, backward design (identifying learning goals prior to designing assessments), and constructive alignment (ensuring teaching methods and assessments directly match those intended goals) within a pragmatic workflow for faculty teams who must first diagnose the current curriculum and only then decide whether, where, and how redesign is warranted. Early versions were drafted and refined through repeated application to typical programme materials (e.g., course syllabi, course learning outcomes, assessment briefs, and programme aims), with successive revisions to improve clarity, reduce ambiguity, and ensure that outputs could be used directly in committee deliberation.

The final prompt set was shaped by two practical requirements. First, each step had to yield outputs that are both actionable (usable for decisions and revisions) and auditable (traceable to inputs and to the reasoning that justified changes). Second, prompts had to support team-based use: the Playbook is intended to structure shared interpretation, negotiation, and documentation, rather than to replace curriculum expertise. For transparency and replication, the full prompt wordings and templates are provided in the Steps and in the Supplement, and the logging and validation safeguards for their use are specified in the Reproducibility and Validation Protocol section.

The Prompt Playbook is grounded in a human-in-the-loop or hybrid-intelligence perspective, in which GenAI augments rather than replaces human expertise (Dellermann et al., 2019). Within this framework, prompts are used to generate alternative perspectives, identify potential gaps, and accelerate information synthesis; however, interpretation, prioritisation, and final decision making remain the responsibility of curriculum leaders, instructors, and other stakeholders. The purpose of the Playbook is

therefore not to automate curriculum redesign, but to support evidence-informed human judgement through structured AI-assisted workflows.

Figure 1 summarises the Prompt Playbook as a three-step workflow in which Step 1 outputs inform Step 2 mapping and Step 3 integration, wrapped by governance/validation procedures and documented through auditable logs.

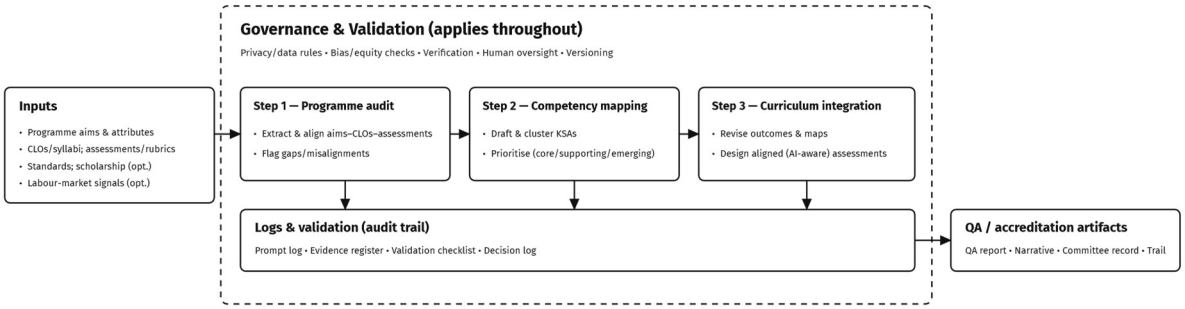

Figure 1. Prompt Playbook workflow

Prompt Playbook workflow

Note: Inputs (programme documents and contextual evidence) flow through a three-step, prompt-supported process: Step 1 audits alignment across aims, learning outcomes, and assessment; Step 2 maps and prioritises KSAs; Step 3 integrates priorities into revised outcomes, curriculum maps, and aligned (AI-aware where relevant) assessments. A governance and validation layer applies throughout, and documentation (prompt log, evidence register, validation checklist, decision log) produces an auditable trail that supports QA and accreditation artifacts.

Step 1: Programme Audit – Reviewing the Curriculum "As Is"

Before introducing new competencies (e.g., data literacy, digital ethics, or responsible use of generative tools), curriculum teams need a clear picture of what is currently taught, how it is assessed, and where misalignments occur. Curriculum reform work emphasises the value of phased change anchored in a shared understanding of the existing curriculum, including how learning outcomes and assessment demands are distributed across courses. Step 1 therefore establishes an evidence-informed baseline: what the programme claims to develop, what students are required to do in practice, and whether assessment tasks provide credible evidence for intended learning outcomes.

Programme Audit is designed for small faculty teams, curriculum committees, or programme directors working with routine programme documentation (syllabi, course learning outcomes, assessment briefs, rubrics, weekly topics, and programme-level aims). Although a single educator can run the prompts, the audit is most effective when outputs are reviewed and revised collaboratively by subject-matter experts. The resulting artifacts are intended to be shareable within departmental processes (planning meetings, QA reviews, and accreditation documentation) and to provide a clear starting point for Step 2 and Step 3. Outputs generated during the audit stage should be interpreted as decision-support artefacts rather than definitive evaluations and should be reviewed by relevant faculty and programme stakeholders.

The personas used throughout the Prompt Playbook serve as cognitive framing devices that encourage the AI system to adopt the perspective of specific stakeholders involved in curriculum development, such as accreditation specialists, labour-market analysts, educational designers, and researchers. Prior work in prompt engineering suggests that role-based prompting can improve output specificity, consistency, and alignment with user goals (White et al., 2024). Within the present framework, personas are intended not to replace expert judgement but to simulate complementary viewpoints that can support human decision-making during curriculum review.

Core Prompts for Programme Audit

Prompts A–D form a lightweight audit that can be completed within an existing planning meeting or departmental retreat. Across the four prompts, teams generate: (i) a concise diagnostic summary, (ii) a prioritised gap list, (iii) a programme map, and (iv) a course learning-outcome table annotated for cognitive demand and priority competencies.

Prompt A: Curriculum health check. Persona: Curriculum-development consultant with expertise in psychology education. Purpose: Review each course syllabus, learning outcomes, weekly topics, and assessment methods to identify strengths, misalignments, redundancies, and likely gaps. Output: A short diagnostic summary and a small set of high-leverage improvement actions (e.g., where assessments fail to evidence specific outcomes, or where topics are repeated without increased cognitive demand).

Prompt B: Mission fit. Persona: Accreditation reviewer. Purpose: Compare programme aims and graduate attributes with course-level learning outcomes and assessment evidence to identify where the programme's stated mission is not consistently implemented across the curriculum. Output: A brief alignment judgement (strengths and risks) and suggested refinements to programme-level aims/attributes, phrased in assessable, discipline-appropriate terms.

Prompt C: Programme map. Persona: Psychology programme director. Purpose: Create a simple course-by-competency matrix indicating where target competencies are introduced, developed, and assessed across the programme. Output: A draft map that makes the distribution and sequencing of competencies visible and highlights areas with limited coverage, uneven progression, or unclear ownership across courses.

Prompt D: Course learning outcome gap finder. Persona: Learning outcomes analyst. Purpose: Tag each course learning outcome by cognitive demand (e.g., using Bloom's taxonomy) and indicate whether it addresses any agreed priority competency area. Output: A table with at least three columns (learning outcome, cognitive level, competency tag) designed to feed directly into Step 3 when learning outcomes and assessments are revised.

Illustrative Example. In the worked example presented in Appendix A, the curriculum audit identified substantial coverage of traditional research methods but limited explicit attention to AI literacy, algorithmic bias, and human–AI collaboration. These findings informed subsequent competency-mapping activities in Step 2.

Using the Results

The outputs from Prompts A–D are designed to be shared with colleagues and used directly in programme-level decision-making. At the programme level, the diagnostic summary and gap list can support QA reporting, accreditation narratives, and

prioritisation of curriculum changes. The programme map provides a visual overview of where competencies are introduced, developed, and assessed, enabling committees to identify sequencing problems (e.g., advanced outcomes assessed before foundational instruction) and to assign responsibility for strengthening coverage.

Traditional curriculum review processes often rely on committee deliberation, manual document analysis, and periodic stakeholder consultation. While these approaches remain essential, they can be time-intensive and may struggle to incorporate rapidly evolving evidence from labour markets, emerging technologies, and policy developments. The Prompt Playbook is intended to complement existing review practices by accelerating information synthesis, generating structured analyses, and surfacing potential curriculum gaps for human evaluation. Its primary contribution is not the replacement of established QA processes, but the provision of a scalable mechanism for supporting evidence-informed curriculum renewal in rapidly changing contexts.

At the course level, Step 1 outputs support targeted revisions to learning outcomes and assessment tasks in Step 3 by clarifying current cognitive demand, identifying mismatches between outcomes and assessment evidence, and highlighting where new or underdeveloped competencies require stronger instruction and assessment. To support transparency and later evaluation, teams should retain audit outputs as time-stamped artifacts and record which documents were used as inputs; the recommended logging, verification, and validation safeguards are set out later in the manuscript under the Reproducibility and Validation Protocol.

Stakeholder Consultation

Although AI-assisted analysis can help identify potential curriculum gaps, curriculum decisions should ultimately be informed through consultation with relevant stakeholders. Students can provide insight into perceived skill gaps and learning needs, employers can highlight emerging workforce expectations, and external partners can offer perspectives on professional standards and future trends. Findings generated through the audit process should therefore be treated as starting points for structured dialogue rather than as final recommendations. Integrating stakeholder consultation throughout the Playbook aligns with established principles of participatory curriculum development and helps ensure that curriculum changes are contextually appropriate and educationally meaningful.

Step 2: Competency Mapping – Using an Illustrative Job-advertisement Protocol

Step 2 of the Prompt Playbook supports programmes in identifying and prioritising high value KSAs in response to GenAI as well as broader social, technological, environmental, and policy shifts. In psychology-related roles, professional guidance and emerging practice literature increasingly point to the relevance of competencies such as data-literate decision-making, responsible use of AI tools, and strengthened ethical and governance awareness in technology-mediated contexts.

Cross-cutting workforce evidence also highlights growing demand for analytical, digital, and human skills (e.g., communication, collaboration, and judgement) across sectors (LinkedIn Corporation, 2023). These trends suggest that curriculum renewal should be informed not only by disciplinary traditions but also by emerging professional expectations. Consequently, psychology programmes require systematic mechanisms for

identifying which AI-related competencies are most relevant, most in demand, and most appropriate for integration into existing learning outcomes.

To address this need, Step 2 introduces a structured competency-mapping process that combines labour-market evidence, scholarly literature, and professional standards to identify priority areas for curriculum enhancement. Rather than assuming a single "future skills" list, Step 2 is designed to help departments make prioritisation decisions transparently and in a way that can be justified to internal and external stakeholders.

This step provides a flexible, replicable protocol for mapping and discussing priority KSAs. It is deliberately topic-agnostic: while examples in this paper emphasise competencies associated with GenAI, the same method can be applied to other curricular priorities (e.g., sustainability, trauma-informed care, or telehealth). Where programmes wish to conduct an empirical scan, the protocol below can be implemented using locally defined sampling and governance decisions.

Like Step 1, the competency-mapping process can be conducted at either the programme level or the individual course level. At the programme level, the focus is on broad graduate capabilities and curriculum coherence. At the course level, the analysis targets specific learning outcomes, assessments, and teaching activities. Whichever level is selected should be maintained consistently across all three steps of the Prompt Playbook.

A Protocol for Labour-Market Mapping (Illustrative Only)

Departments may choose to use job advertisements, professional/accreditation standards, and selected scholarly trends as one input into competency mapping. Because job-posting data can be partial and biased by sector, geography, and recruitment practices, outputs should be treated as discussion prompts and triangulated with disciplinary standards and local programme aims.

Triangulation rule (illustrative). No KSA becomes a core curriculum priority unless supported by ≥2 of: disciplinary standards, programme mission/graduate attributes, feedback from faculty, employers, alumni, students, and professional bodies, labour-market scan, or scholarly synthesis; single-source KSAs are treated as provisional and reviewed in later cycles.

Minimum viable scan. Sample 40–60 relevant postings across 4–6 graduate job families from 2–3 sources, within 6–12 months, in the programme's primary graduate geography; de-duplicate and document the sampling frame and coding rules.

The workflow below describes a practical procedure that departments can adapt locally.

In practice, a curriculum team may identify a bounded set of plausible graduate job families and sample a manageable number of recent advertisements from relevant public sources. The aim is not to construct a representative labour-market dataset, but to surface recurring requirements that can be triangulated with disciplinary standards and local programme aims. Using the text of these advertisements, especially essential/desirable criteria and role responsibilities, the team then groups recurring requirements into broad KSA clusters meaningful for curriculum review. Because this step is interpretive, coding rules and category labels should be documented transparently, whether the work is carried out manually or with tool support, so that faculty can inspect, challenge, and refine the resulting priorities.

To improve comparability over time and reduce ad hoc category drift, departments may optionally align clusters to established skills/occupation taxonomies (e.g., ESCO or O*NET; European Commission, 2025; U.S. Department of Labor, Employment and Training Administration, 2025) or to existing departmental competency models, then record any adaptations made to fit local regulatory and disciplinary requirements.

Outputs should not be used to "teach to the job advert"; rather, they are intended to support deliberation about sequencing, assessment evidence, and where emerging competencies can be integrated without displacing foundational disciplinary learning.

Core Prompts for Competency Mapping

Prompts E–H translate external labour-market, scholarly, and professional signals into curriculum-relevant insight. They may be used sequentially or selectively, depending on departmental goals and available resources. Where teams do not conduct a formal labour-market scan, the prompts can still be used to generate a structured qualitative synthesis (e.g., recurring knowledge, skills, and abilities with illustrative examples) and to triangulate priorities using scholarship and relevant professional standards.

Prompt E: Job-Market Miner. Persona: Labour-market analyst. Task: Identify frequently requested KSAs in a defined sample of recent job postings relevant to the programme and a specified theme (e.g., psychology and AI), drawing on one or more public sources and a defined sampling window. Output: A table of priority KSAs with either (a) counts and representative excerpts (for small local samples) or (b) frequencies (for larger samples), alongside example job titles and representative phrasing from postings. The output also records the sampling frame (sources, dates, inclusion criteria) to support later interpretation and replication.

Prompt F: Scholarship Mapper. Persona: Research synthesis specialist. Task: For each priority KSA, identify and briefly summarise scholarship, professional standards, or domain guidelines that provide a rationale for inclusion in the curriculum. Output: An annotated list mapping each KSA to at least one scholarly source or standard. For example, ethical use of AI may be linked to scholarship on responsible or trustworthy AI, the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2017), and UNESCO guidance on GenAI in education and research (UNESCO, 2023). Note: Any sources suggested by the tool are verified against the original documents before citation.

Prompt G: Strategic KSA Forecaster. Persona: Educational futurist or organisational psychologist. Task: Identify KSAs likely to increase in importance over the next three to five years and justify why, drawing on observable technological, demographic, regulatory, and policy trends. Output: A concise set of anticipated shifts in skill demand (e.g., five), including assumptions and uncertainties so that forecasts can be revisited as conditions change.

Prompt H: Curriculum Gap Analyser. Persona: Accreditation advisor. Task: Compare the priority KSAs identified through Prompts E and G (and, where relevant, triangulated via Prompt F) with the programme's existing course learning outcomes (CLOs) and assessment evidence. Output: A matrix indicating which KSAs are currently addressed (and at what level), which are partially covered, and where integration is needed, highlighting gaps, redundancies, and sequencing opportunities across the programme.

The matrix distinguishes between competencies that are taught and those that are assessed with credible evidence, as these may not align.

Illustrative Gap-analysis Matrix

To make the illustrative protocol more concrete, Table 1 presents a schematic example of the gap-analysis matrix generated via Prompt H. The purpose is to show the structure of the decision aid – coverage, priority, and integration options – rather than to report findings from an empirical labour-market study. To reflect the distinction emphasised in Prompt H, the schematic separates whether a KSA is taught/explicit in outcomes from whether it is assessed with credible evidence, as these do not always align.

Table 1. Example Gap-Analysis Matrix (Schematic)

KSA (illustrative)	Taught / explicit in CLOs	Assessed with credible evidence	Priority	Recommended integration
Programming literacy (e.g., Python/R, reproducible analysis workflows)	No	No	High	Core option (methods/statistics pathway)
Prompting and human–AI interaction (prompt design, verification, documentation)	No	No	High	Core (embedded across courses)
Critical thinking	Yes	Partial	High	Strengthen assessment evidence (already embedded)
Data analysis	Partial	Partial	High	Core (required); strengthen sequencing and assessment
Ethical use of AI and bias awareness	Partial	Partial	High	Core (required); integrate across methods/ethics/applied modules
Communication	Yes	Yes	High	Maintain (already embedded)
Continuous learning and professional adaptability	Partial	Partial	High	Core (embedded across courses; strengthen reflective evidence)
Human–AI collaboration (role allocation, oversight, risk boundaries)	No	No	High	Strand, capstone, or placement-based integration

Note: KSA labels, coverage ratings, priority levels, and integration decisions should be adapted to local programme aims, staff capacity, student preparation, national contexts, and accreditation requirements. “Programming literacy” may be operationalised as coding, scripting, tool-mediated reproducible analysis, or statistical workflow fluency depending on programme identity and resourcing. In practice, departments may use finer-grained scales (e.g., introduced/developed/mastered) in place of yes/partial/no.

In practice, teams use the matrix to make sequencing decisions (what belongs in core courses versus strands, electives, or capstones) and to identify where learning outcomes and assessment evidence require strengthening. Because psychology curricula are

typically shaped by disciplinary guidance as well as external signals, Step 2 can also triangulate KSA clusters against recognised psychology education frameworks. The American Psychological Association (APA) Guidelines 3.0 for the Undergraduate Psychology Major (2023) specify five overarching goals and associated learning outcomes that can be used to justify and document curricular priorities in discipline-specific terms. Table 2 provides an illustrative mapping of common KSA clusters to those goals. In European contexts, programmes may additionally align mapped competencies to EuroPsy/EFPA standards and competence expectations as appropriate (European Federation of Psychologists' Associations [EFPA], 2025).

Table 2. Illustrative Mapping of KSA Clusters to APA Guidelines 3.0 Goals (Psychology)

KSA cluster (illustrative)	**Examples of competencies / KSAs (adapt locally)**	**APA Guidelines 3.0 goal alignment (examples)**
Psychological content knowledge and application	Core concepts; major subfields; integrative themes; applying psychology to real problems	Goal 1: Content Knowledge and Applications (e.g., 1.1–1.5)
Scientific inquiry, research methods, and data literacy	Research design; measurement; evidence evaluation; statistical reasoning; reproducible analysis habits	Goal 2: Scientific Inquiry and Critical Thinking (e.g., 2.1–2.4)
Values-based and ethical reasoning (including technology ethics)	Ethics in research/practice; fairness and bias awareness; harm reduction; responsible tool use and disclosure	Goal 3: Values in Psychological Science (e.g., 3.1–3.3) and links to Goal 5 (professional judgement)
Sociocultural and intercultural responsiveness	Intercultural competence; inclusion; culturally informed inquiry and practice; community engagement	Goal 3 (e.g., 3.2–3.3) and Goal 2 (sociocultural factors in research practices; 2.3)
Communication and psychological literacy	Written/oral communication; audience adaptation; psychological literacy; evidence-based argumentation	Goal 4: Communication, Psychological Literacy, and Technology Skills (e.g., 4.1–4.4)
Technology skills and human–AI interaction	Verification/checking; prompting practices; documentation; human oversight; appropriate use boundaries	Goal 4 (technology skills; 4.4) and Goal 5 (workforce preparation; e.g., 5.5)
Personal and professional development (lifelong learning)	Self-regulation; project management; teamwork/collaboration; professional judgement; continuous learning	Goal 5: Personal and Professional Development (e.g., 5.1–5.6)

Note. This mapping is illustrative and intended to support curriculum discussion and documentation. Departments should adapt cluster definitions and goal links to local programme identity, staffing, and national accreditation requirements.

This triangulation provides a defensible rationale for deciding which KSAs become core, strand-based, or capstone-level targets in the Curriculum Integration step.

Using the Results

The outputs of Step 2 are inputs to structured curriculum deliberation rather than final decisions. Prompt E identifies candidate knowledge, skills, and abilities from external signals; Prompt F triangulates these candidates against scholarship and relevant professional standards; Prompt G supports forward-looking prioritisation; and Prompt H consolidates these inputs into an actionable gap-analysis matrix. Table 2 illustrates how priority KSA clusters can be documented in explicitly psychology-relevant terms using the APA Guidelines for the Undergraduate Psychology Major, strengthening the disciplinary rationale for decisions about what belongs in core courses versus electives, capstones, placements, or co-curricular activities.

In practice, Step 2 helps teams (a) prioritise a manageable set of KSAs that are both professionally relevant and educationally meaningful, (b) justify those priorities through transparent triangulation with scholarship, ethics, and disciplinary standards, and (c) document where current provision appears strong, partial, or absent. These outputs are not redesign decisions in themselves. Rather, they function as review-stage artifacts that support structured faculty deliberation about whether curricular change is needed, where it is needed, and what form it should take. Only after this review and prioritisation stage do teams move, if appropriate, to Step 3.

Because gap analysis is partly interpretive, competing readings are expected rather than treated as procedural failure. For example, one group may judge a KSA as “partially covered” while another sees it as adequately embedded but weakly assessed. In such cases, the matrix functions as a deliberative aid: teams revisit the coding rules, inspect the underlying CLOs and assessment evidence, and resolve differences through documented committee judgement rather than by treating the generated categorisation as definitive.

Step 3: Curriculum Integration – Designing Competency-Based, AI-Aware, Inclusive Learning

Step 3 supports both programme-wide curriculum redesign and course-specific enhancement activities. With priority competencies identified and justified in Step 2, the final stage of the Prompt Playbook supports faculty in translating selected priorities into concrete curriculum assets: revised learning outcomes, inclusive learning activities, and aligned assessments. This is the point at which the framework moves from diagnosis to design. In other words, Step 3 is distinct from audit proper: whereas Steps 1–2 are concerned with reviewing, mapping, and interpreting the curriculum as it currently exists, Step 3 addresses what departments may choose to do in response to those validated findings.

Consistent with backward design, Step 3 begins by refining learning outcomes and then aligning teaching and assessment accordingly (Bruff, 2019; Wiggins & McTighe, 2005). Its function is developmental rather than diagnostic. The emphasis is on strengthening the priority competencies surfaced through Step 2 – such as data literacy, ethical reasoning, and discipline-specific professional skills – while making expectations about appropriate use of GenAI explicit and assessable. At this stage, faculty judgement and disciplinary expertise are central: generative tools can support drafting and option generation, but decisions about what counts as evidence of learning, which competencies warrant explicit integration, and what must be demonstrated without tool assistance, remain human responsibilities. Where a generated suggestion appears pedagogically weak, misaligned

with disciplinary expectations, or in tension with programme identity, the appropriate response is revision or rejection – not accommodation to the tool.

Core Prompts for Curriculum Integration

Prompts I–K guide design decisions that translate the priorities identified in Step 2 into assessable curriculum assets. They are intended to support disciplined drafting and alignment (learning outcomes → learning activities → assessment evidence) while keeping faculty judgement, disciplinary standards, and ethical constraints central.

Prompt I: CLO Generator. Persona: Curriculum designer. Task: Revise CLOs to reflect the priority KSAs identified in Step 2, using observable verbs (e.g., Bloom's taxonomy) and tagging each outcome to relevant disciplinary or accreditation standards where applicable (e.g., APA Guidelines for the Undergraduate Psychology Major). Output: A revised set of CLOs, each tagged with an indicative cognitive level and an aligned standard. An illustrative outcome is: "Students will critique AI-generated psychological claims in news media, identify methodological flaws, and propose evidence-based alternatives," tagged at an evaluative level and linked to an APA scientific thinking goal. Depending on programme priorities, outcomes may instead foreground competencies such as teamwork, intercultural communication, sustainability literacy, or professional ethics.

Prompt J: SMART Criteria Check. Persona: QA advisor. Task: Review existing or newly drafted CLOs for clarity and assessability using SMART criteria (specific, measurable, achievable, relevant, and – where appropriate – time-bounded), and revise where needed. Output: A revised set of CLOs with brief annotations explaining how each meets the criteria and how it links to assessment evidence. For example, an outcome such as "Students will understand AI ethics" can be revised to: "By the end of the module, students will evaluate three ethics case studies using a published rubric and identify at least two fairness risks and two mitigation strategies for each case." The accompanying note clarifies how the revised outcome is specific and measurable and how it will be evidenced through the assessment, supporting transparency for learners and staff (Panadero et al., 2023). Teams can calibrate the level of specificity to local QA conventions.

Prompt K: AI-Aware Activities and Assessments. Persona: Inclusive learning designer. Task: Design a pair of complementary tasks – one AI-supported and one AI-restricted – that target the same learning outcome. The AI-restricted task is designed to elicit unaided reasoning, disciplinary judgement, and/or ethical reflection; the AI-supported task develops capability in evaluating, using, and documenting tool-assisted work. Output: A short activity or assessment pair with explicit disclosure expectations (what students must report about tool use) and clear marking criteria. For example, in a biopsychology course, students might use a large language model to draft a summary of a neuroscience article and then critique the accuracy and evidential basis of that summary; in the paired AI-restricted task, students critique the same article independently without access to AI-generated text. Both tasks align to an outcome such as: "Evaluate the reliability of neuroscience claims using primary sources and appropriate reasoning." Activities are designed and implemented in line with the ethical, equity, and data-governance safeguards set out in the Ethics and Data Governance section.

Illustrative Example. The identified competency gaps led to the development of revised learning outcomes and an AI-bias auditing activity that aligned with existing APA competency domains while introducing new AI-related capabilities.

Design Priorities and Flexibility

The goal of Step 3 is not to replace existing content, but to embed priority KSAs in ways that preserve disciplinary integrity and improve alignment between outcomes and assessment evidence. For example, ethical reasoning about tool use can be developed through case-based work in clinical, research, or organisational psychology; data literacy can be strengthened in methods and statistics through reproducible workflows and analysis of authentic datasets; and intercultural responsiveness can be developed through group projects, community-engaged learning, or placement-related reflection. Departments may decide that some competencies should be woven into core courses, while others are more appropriately developed in strands/pathways, electives, capstones, placements, or co-curricular workshops. Flexibility is essential, particularly where staff expertise, infrastructure, or student preparation varies. Proposed curriculum changes should be reviewed through appropriate stakeholder consultation processes prior to implementation.

Examples in Context: Psychology Courses

Illustrative examples of Step 3 are embedded throughout the Playbook. One biopsychology example asks students to evaluate tool-generated claims about neural mechanisms and to produce guided critiques using paired AI-supported and AI-restricted tasks. A psychological measurement mini-module has students compare tool-assisted and manual scoring approaches to explore reliability, validity, and error. These examples are intended to show how responsible integration can reinforce – rather than displace – core competencies such as research literacy, ethical judgement, and disciplinary critique. Parallel designs can be developed in other disciplines by pairing AI-supported and AI-restricted tasks that target the same outcome while making evidence and disclosure requirements explicit.

Lessons Learned from the Illustrative Application

Applying the Prompt Playbook to the illustrative curriculum yielded three practical insights. First, AI-assisted auditing rapidly identified areas of curriculum omission that might have been overlooked during manual review. Second, labour-market scanning generated useful discussion points for curriculum committees, although human interpretation remained essential when judging the relevance and educational value of identified competencies. Third, the process highlighted the importance of balancing AI-related skills with enduring disciplinary competencies rather than treating them as competing priorities. These observations support the view that the Playbook functions most effectively as a structured decision-support framework embedded within existing academic governance processes.

Using the Results

The CLOs, activities, and assessments produced in Step 3 are designed to feed directly into course outlines, assessment maps, and programme-level reviews. Outputs can be used in curriculum committee discussions to check alignment and progression, included as artifacts in accreditation documentation, and adapted for staff development as departments build shared capacity for AI-aware teaching and assessment. Because

prompts can be reused and iterated over time, departments can update courses gradually rather than relying on large-scale overhauls.

Taken together, Prompts I–K support departments in translating priority KSAs into clear, assessable learning outcomes and aligned learning activities and assessments. Prompt L (student-facing guidance on acceptable tool use – see below) then operationalises these decisions for learners by clarifying expectations, disclosure requirements, and boundaries. Combined with the audit and mapping work in Steps 1 and 2, Step 3 provides a practical route for strengthening coherence, evidentiary validity, and ethical governance as professional expectations and educational technologies evolve.

Ethics and Data Governance

The Prompt Playbook is designed to support responsible curriculum development in a period of accelerating technological change, and the ethical considerations in this section apply across all three steps of the framework. GenAI tools can support drafting, synthesis, and formative feedback, but they also introduce material risks for psychology education, including biased outputs, misinformation, automation bias (over-trust in system suggestions), academic integrity failures, privacy breaches, and unequal access. These risks shape how prompts should be written, how outputs should be interpreted, and how students experience learning and assessment in technology-mediated environments.

Prior work indicates that large language models can reproduce or amplify harmful stereotypes and structural biases (Bender et al., 2021) and that users may exhibit automation bias even when system outputs are flawed (Burton et al., 2020; Dratsch et al., 2023). These concerns align with professional ethics expectations in psychology, including the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2017), and with broader AI ethics principles emphasising fairness, accountability, transparency, and harm reduction (Fjeld et al., 2020; Floridi & Cowls, 2019; Tsamados et al., 2022). In educational settings, risks can arise when students treat tool-generated case formulations, interpretations, or statistical explanations as authoritative without verification. The Playbook is designed to counter these risks by requiring verification-oriented prompting, documenting assumptions and evidence, and pairing AI-supported work with AI-restricted tasks and structured reflection.

The framework also recognises that inequities in access to tools – arising from bandwidth constraints, licensing costs, language barriers, interface complexity, or disability-related exclusions – can deepen educational disparities if left unaddressed. Accordingly, the Playbook is tool-agnostic and includes low-tech variants of all prompts; it encourages instructors to provide inclusive alternatives that do not assume proficiency with specific platforms or paid features. This is particularly important given evidence of uneven AI literacy and access among students (Bećirović et al., 2025; Sullivan et al., 2024). Departments are encouraged to incorporate universal design and digital equity considerations into assessment planning (Rose & Meyer, 2007; UNESCO, 2023).

The framework addresses ethical concerns in four interlocking ways.

Human-in-the-loop design. Prompts are intended to support faculty deliberation rather than automate curriculum decisions. Generative tools function as drafting and reflection aids, while faculty remain responsible for interpreting outputs, verifying claims, and aligning curricular choices with disciplinary standards, accreditation expectations, and institutional requirements. Where generated outputs conflict with disciplinary expertise,

local programme aims, or committee judgement, the tool output has no presumptive authority: it is treated as a prompt for further review, not as a basis for override. This orientation is consistent with hybrid intelligence approaches that combine human and machine capabilities while keeping human judgement and accountability central (Dellermann et al., 2019).

Transparent and auditable use of outputs. Prompt outputs are treated as provisional artifacts (options and rationales), not prescriptions. Teams are encouraged to document which prompts were used, what inputs were provided, what changes were made, and how human judgement shaped final decisions. This documentation supports internal QA and provides an evidence trail for external review. The Reproducibility and Validation Protocol (below) specifies minimal expectations for logging and verification.

Data governance in prompt use. When curriculum work involves course artifacts, student work, or externally collected text (e.g., job advertisements), departments should follow institutional data-protection requirements and relevant guidance on data ethics and governance (Boddington, 2017; Xafis et al., 2019). Identifiable student information should not be entered into commercial tools unless explicitly permitted by institutional policy; anonymised, aggregated, or synthetic materials can be used instead. Where feasible, departments may prefer institutionally hosted solutions to support compliance with applicable privacy and data protection obligations, including the General Data Protection Regulation (European Parliament & Council of the European Union, 2016).

Bias mitigation and ethical reflection. Learning activities and assessments involving generative tools incorporate critique, comparison, and metacognitive reflection. Students are guided to identify inaccuracies, examine potential harms, and articulate ethical boundaries of tool use within psychological science and practice. In Step 3, this is implemented through paired AI-supported and AI-restricted tasks that target the same learning outcomes while requiring evidence, justification, and disclosure.

Sidebar: Student-Facing Policies on Acceptable AI Use

Where AI-supported learning activities or assessments are introduced, expectations should be made explicit through course-level policy language. At minimum, policies should specify: which tools (if any) are permitted; what students must disclose about tool use; what constitutes unacceptable conduct (e.g., undisclosed ghost-writing, misrepresentation of authorship, fabricated references); and where students can seek support if tool outputs cause confusion or distress. Policies should align with institutional regulations and academic integrity frameworks and should be discussed explicitly in class to reduce ambiguity and promote ethical, transparent practice. Sample wording can be generated and adapted using Prompt L.

Prompt L: AI policy generator. Persona: Course coordinator. Task: Draft a student-facing policy paragraph for a course syllabus that defines permitted uses, disclosure requirements, and academic integrity expectations. Output: Approximately 150 words of policy text suitable for a syllabus, with optional variants (e.g., “AI-restricted assessment,” “AI-supported drafting allowed with disclosure,” “no external tools permitted”).

The Prompt Playbook does not promote uncritical adoption of generative tools. Instead, it treats GenAI as a catalyst for disciplined curriculum review and redesign, with ethics, data governance, accessibility, and faculty oversight embedded in the workflow rather than added as afterthoughts. In this sense, responsible tool use is positioned as one

component of a broader strategy for maintaining curricular coherence, disciplinary integrity, and accountability under changing professional and technological conditions.

Reproducibility and Validation Protocol

To operationalise the governance principles outlined in the Ethics and Data Governance section, this subsection specifies a minimal, auditable protocol for producing, checking, and documenting outputs generated using the Prompt Playbook (Prompts A–L) across Steps 1–3. The protocol treats outputs from generative tools as decision-support artifacts – inputs to faculty deliberation – rather than authoritative determinations; final curricular judgements remain with programme governance bodies.

Here, reproducibility should be understood in three related but distinct senses. First, procedural reproducibility refers to whether another team could rerun the documented workflow using the same evidence pack, prompt version, and tooling conditions and inspect how outputs were produced. Second, output stability refers to whether materially similar judgments, categories, or recommendations recur across repeated runs, sessions, or models, even if wording differs. Third, decision robustness refers to whether any variation in generated output changes the human curriculum decision that follows. Exact textual replication is not assumed and, in many LLM environments, should not be expected.

Protocol elements (required unless stated otherwise). First, the team defines the scope of use for the review cycle (e.g., programme-wide audit versus selected courses) and records a brief boundary statement specifying what the tool use will not include (e.g., no student-identifiable data; no creation of new empirical claims; no unverified citations; no institution-identifying information in prompts). Second, the team compiles a Curriculum Evidence Pack to be used consistently across runs, including current syllabi, course learning outcomes, assessment briefs and rubrics, credit structures, prerequisites, programme learning outcomes, accreditation and disciplinary requirements, and any external reference set used in Step 2 (e.g., selected professional guidance or labour-market samples). This ensures that outputs remain anchored in traceable inputs and that teams can distinguish between evidence-based inferences and speculative suggestions.

Third, each run is captured in a Run Record and Prompt Log to enable replication and later audit. A template for this documentation is provided in Table 3 in Appendix A. At minimum, the log records: the prompt identifier (A–L) and local version (with a brief rationale for any edits); operator role; date/time; tool/provider and model/version as reported at runtime; any accessible generation settings (e.g., temperature/top-p) and whether any retrieval/browsing mode was enabled; the exact prompt text; the specific input documents or excerpts supplied; and the output captured verbatim (or stored in an appendix/supplement). Because model behaviour can vary across releases, outputs should be treated as time-stamped artifacts tied to a specific tool and model/version rather than assumed stable findings.

Fourth, repeat-run check for high-stakes uses (recommended; required where outputs are likely to shape programme-level decisions). For prompts that materially influence competency prioritisation, gap diagnosis, or assessment redesign, teams should repeat the run at least once under the same documented conditions, ideally in a separate session. Where feasible, a second check may also be conducted using an alternative model or provider. The aim is not to demand identical prose, but to examine whether the

same substantive issues, KSA priorities, alignment concerns, or recommended actions recur. Stable convergence across runs increases confidence; substantial divergence indicates that outputs should be treated as exploratory only and subjected to deeper human review.

Fifth, outputs undergo a structured Validation Checklist before they are used in programme decisions or documentation. Where feasible, validation is performed by a reviewer other than the initial operator. The checklist includes three compulsory checks: factual integrity (no fabricated standards, policies, or references; any suggested citations are verified against original sources before inclusion), alignment integrity (learning outcomes use observable verbs and are level-appropriate; learning-outcome-to-assessment links are explicit; mapping decisions use consistent definitions and coding rules), and completeness/specificity (assumptions and missing information are flagged rather than filled; recommendations specify what changes, where, and how it will be evidenced). Items that fail are classified as critical defects (must be corrected before use) or minor defects (may be corrected during committee refinement).

Sixth, when reviewers or committee members disagree with a generated output, or with one another's interpretation of that output, the item is not advanced directly into curriculum change. Instead, the team returns to the underlying evidence pack and asks three questions: (a) Is the suggestion supported by the supplied curriculum evidence? (b) Is it consistent with disciplinary standards, accreditation requirements, and programme aims? (c) Would adopting it improve alignment or merely reflect the tool's generic preferences? Items that remain contested after this check are either revised manually, flagged for committee discussion, or rejected. In all such cases, the Decision Log records the disagreement, the grounds for resolution, and the final governance outcome.

Seventh, reviewers also record whether outputs are stable at the level of interpretation. A simple three-part notation may be used: (a) stable – core categories/recommendations recur across runs; (b) partially stable – core themes recur but ranking, phrasing, or emphasis varies; (c) unstable – different runs produce materially different priorities, mappings, or recommendations. Outputs classified as unstable should not be used directly in programme documentation or redesign decisions without additional triangulation, manual coding, or committee review.

Eighth, the team conducts an explicit equity and bias review consistent with the safeguards described in the preceding section, with particular attention to Step 2 artifacts that draw on external signals (e.g., job advertisements or skills taxonomies). The review records sampling boundaries and coverage (sector, geography, seniority, role types), potential sources of representational bias, and mitigations adopted (e.g., triangulation with disciplinary standards, accessible assessment alternatives, scaffolded supports). Where AI-supported assessment redesign is generated in Step 3, the review also checks that proposed requirements do not introduce inequitable access burdens (e.g., tool availability, disclosure demands, workload asymmetries).

Ninth, curricular decisions are captured in a Decision Log that links validated outputs to human governance actions: which recommendations were accepted, revised, or rejected; the reason; the approving body/meeting; and the resulting changes to learning outcomes, assessments, or course-level policy language (including effective dates). The Prompt Log, Validation Checklist, and Decision Log are stored in the programme's QA repository with

appropriate access controls, enabling internal accountability and supporting external review where appropriate.

Tenth, as an optional reliability and stability check for mapping activities, departments may double-code a sample (e.g., 20–30% of courses) with two independent raters and reconcile discrepancies using a short coding guide. Where tools are used directly in the mapping process, teams may also compare a small sample of repeat runs across sessions, and, where feasible, across models, to assess whether the same mapping decisions recur. Simple agreement reporting can be used internally to indicate both inter-rater reliability and mapping stability over time.

To make this procedure concrete, the supplementary material includes one worked example based on an actual captured run using a Prompt Playbook prompt and a bounded evidence pack. The example reproduces the prompt, tooling context, and verbatim output, then shows how the Validation Checklist and Decision Log are applied (see Table 4 in Appendix A) to classify defects, identify usable elements, and document which suggestions were accepted, revised, or rejected.

Adopting this protocol enables Prompt Playbook outputs to be interpreted as documented, verifiable inputs to curriculum deliberation – supporting transparency, reproducibility, and defensible decision-making. The next subsection outlines pragmatic options for integrating these documentation and validation activities within existing curriculum review cycles and faculty capacity constraints.

Implementation Considerations

The Prompt Playbook is designed to fit within existing departmental and institutional processes rather than prescribe a single implementation pathway. This section outlines pragmatic ways instructors and programme teams can adopt the Playbook at different levels of scope and intensity while maintaining ethical, inclusive, and academically defensible practice.

Use Cases and Scope of Use

The Playbook can be used by individual instructors, small course teams, curriculum committees, or whole departments. At the course level, an individual lecturer might use Step 1 prompts (e.g., Prompt A for a curriculum health check and Prompt D for learning outcome tagging), then apply selected Step 3 prompts (I–K) to revise CLOs and align assessments and use Prompt L to draft a student-facing policy statement on acceptable tool use. At the programme level, a committee may distribute prompts across members (e.g., Prompts B–C for mission fit and programme mapping; Prompts E–H for competency mapping), then synthesise outputs in a joint meeting to inform decisions about sequencing, course ownership of competencies, and documentation for QA. At the department level, the method can be applied across year levels or sub-disciplines, with optional locally defined prompts (e.g., student feedback synthesis, accreditation cross-checks, inclusion audits) added as needed.

Across these use cases, the framework supports partial and phased adoption. A programme may begin with Step 1 only, pilot the approach in one year level, or trial one assessment redesign before scaling. This staged adoption matters conceptually as well as practically: departments may use the Playbook solely for audit and mapping purposes within QA or accreditation cycles, without proceeding to redesign. Step 3 is therefore best

understood as an optional extension of the method rather than as a defining component of curriculum audit itself. Full adoption is not required for meaningful improvement, and although examples in this paper are drawn from undergraduate psychology, the same patterns of individual, team, and department-level use are transferable across disciplines.

Curriculum Review Cycles and Faculty Capacity

Most departments will need to implement curriculum renewal in a staged way, aligning effort with existing QA cycles and realistic workload. A light-touch annual pass can be used to refresh a small set of outcomes and assessments in targeted courses, while fuller use of the Playbook can be reserved for formal review cycles (e.g., accreditation- or QA-linked reviews). This approach is consistent with calls for ongoing, alignment-focused curriculum review in professional education contexts (Kulasegaram et al., 2018). To reduce burden, prompts are designed for time-efficient use in short workshops or retreats that focus on one course, one assessment, or one alignment issue at a time; teaching-and-learning centres can also adopt the prompts as staff-development resources or mentoring tools (Costa et al., 2024; Mah & Groß, 2024). Where departments adopt the Playbook at scale, the Reproducibility and Validation Protocol provides a minimal structure for logging, verification, and decision documentation without requiring extensive additional infrastructure.

Equity, Resources, and Access

Institutional access to generative tools varies widely. The Playbook does not assume access to commercial platforms: instructors can work with pre-filled prompts, locally generated summaries, example outputs, or manual processes that simulate the same structured reasoning. Where tools are used, implementation should comply with institutional data governance and accessibility requirements. To support digital equity, Step 3 encourages paired AI-supported and AI-restricted variants of activities and assessments so that students are not disadvantaged by lack of access, disability-related barriers, language constraints, or ethical reservations, consistent with emerging guidance and evidence on unequal access and literacy (Bećirović et al., 2025; Sullivan et al., 2024; UNESCO, 2023). Teams are also encouraged to surface resource constraints explicitly – such as uneven access to instructional design support, teaching assistants, or workload buy-out – and to consider how these constraints shape what is feasible and who can participate in curriculum renewal.

Process Transparency and Documentation

A practical advantage of the Playbook is that it makes curriculum work visible. Decisions about learning outcomes, assessment evidence, and the integration of emerging competencies often occur informally; prompt-guided workflows generate artifacts that can be shared with colleagues, used in moderation and review, and revisited in subsequent cycles. Teams are encouraged to retain time-stamped outputs and to document (i) which prompts were used, (ii) what inputs were provided, (iii) what changes were adopted or rejected, and (iv) the rationale for decisions. This supports institutional memory and continuous improvement and provides transparent evidence of deliberation and oversight for QA and accreditation. Where disagreement arises, transparency matters as much as the final decision. Departments should therefore record not only what was adopted, but also where the tool's framing was judged misleading, over general, or

incompatible with local disciplinary priorities, so that later review cycles can see how human judgement shaped the final curriculum outcome.

Proposed Pilot Study and Future Research

To enable future empirical evaluation of the Prompt Playbook, a small-scale mixed-methods pilot is proposed to assess feasibility, reliability (mapping stability), and proximal effects on curriculum alignment in undergraduate psychology. A feasible pre–post curriculum-mapping design would involve two second-year courses with contrasting assessment formats (e.g., Biopsychology and Research Methods II). Step 1 (Prompts A–D) and Step 2 (Prompts E–H) establish an audit baseline and prioritise KSAs, followed by Step 3 (Prompts I–K) to revise CLOs and assessment evidence, with Prompt L clarifying student-facing expectations. Primary measures include feasibility (time-on-task, number of prompt/revision iterations, brief workload/usability Likert ratings); reliability and stability (independent double-coding of a subset of CLO→KSA mappings; Cohen's κ; and a small repeat-run comparison to assess whether core mapping judgements recur across sessions and, where feasible, across models); and proximal educational value (expert panel ratings of CLO–assessment alignment pre/post using a short rubric). Any illustrative student work samples should be anonymised and managed under institutional ethics and data governance. Findings will be used to refine prompts and facilitation guidance and to inform programme-level and multi-institutional studies.

The following section summarises limitations and directions for future evaluation.

Limitations and Future Research Directions

This paper presents a methodological framework for curriculum audit and competency mapping, together with an optional follow-up design component under conditions of changing professional expectations and expanding use of GenAI in higher education. It does not present an empirical evaluation of the framework in use. The Prompt Playbook is intended to scaffold curriculum review and programme design by generating auditable decision-support artifacts; it does not prescribe fixed learning designs, validated learning outcomes, or universal competency sets.

Several limitations should be noted. First, the framework and its procedural tools – including the illustrative competency-mapping protocol and schematic gap-analysis matrix – have not yet been tested systematically across institutions, cohorts, or disciplines. Most examples in this paper remain illustrative rather than derived from a formal empirical dataset, and evidence of impact on learning, assessment quality, or faculty practice remains to be established. To increase methodological transparency, however, Appendix A includes a supplementary worked example based on one actual captured LLM run, used solely to demonstrate how imperfect outputs are documented and reviewed under the validation protocol.

Second, although the Playbook is designed to be domain-flexible, its worked examples are oriented to psychology curricula and primarily English-language materials. Adaptation and validation are therefore needed in diverse disciplinary, cultural, and regulatory contexts, including settings where professional standards, curriculum structures, and acceptable tool use differ. Third, competency signals used in Step 2 (e.g., job advertisements or skills taxonomies) are inherently partial and may reflect sectoral, geographic, or recruitment biases. As a result, Step 2 outputs require triangulation with disciplinary standards and programme aims, and they should be treated as inputs to

deliberation rather than direct prescriptions. Fourth, the technological landscape is dynamic: tool capabilities, access conditions, model behaviour, and institutional policies change quickly. Even when prompts and source materials are held constant, outputs may vary across sessions, providers, or model releases. The Playbook therefore cannot assume exact textual reproducibility; instead, it depends on documented procedures, validation checks, and explicit review of whether substantive curriculum judgements remain stable enough to support defensible decisions. Finally, implementation conditions vary widely. The framework assumes access to baseline documentation and the capacity to run or simulate the prompts, but not all faculty will have consistent access to tools, workload support, or local policy clarity on data governance and academic integrity. In this framework, audit and mapping functions are primary, whereas redesign is treated as a subsequent, human-governed application of findings generated through those earlier stages.

These limitations suggest several directions for future research and practice. A priority is feasibility and process evaluation: pilot studies can examine how faculty teams use the prompts in authentic curriculum settings, how outputs shape deliberation, and which barriers and enablers affect adoption (e.g., workload, facilitation, policy clarity, and disciplinary consensus). A second direction is educational impact: studies can investigate how paired AI-supported and AI-restricted assessments influence learning, engagement, metacognition, and academic integrity, including whether assessment evidence better matches intended learning outcomes. A third direction concerns governance and equity in practice: research can examine how departments operationalise disclosure expectations, manage differential access and literacy, and implement privacy safeguards when curricula incorporate tool-mediated work. A fourth direction is methodological: further work is needed on the validity and interpretability of labour-market-derived competency signals and on best practices for triangulating such signals with disciplinary standards and local programme identities.

To support transparent replication and cumulative improvement, future pilots should test and refine the Reproducibility and Validation Protocol by examining output stability across repeated runs, sessions, and models; identifying which prompt types are most sensitive to variation; and determining what level of stability is sufficient for different curriculum uses (e.g., exploratory brainstorming, mapping support, or programme-level QA documentation).

Conclusion

This paper introduced the Prompt Playbook, a flexible and ethically grounded method for curriculum audit in psychology under conditions of changing professional expectations and expanding use of GenAI in higher education. The Playbook links three stages – programme audit, illustrative competency mapping, and an optional curriculum-integration phase – so that departments can first make gaps and priorities visible and then, where appropriate, translate validated findings into revised learning outcomes, learning activities, and assessment evidence through transparent, collaborative processes.

To support trustworthy adaptation, the Playbook pairs its prompt set with explicit safeguards for ethics, equity, data governance, and verification, alongside a minimal reproducibility/validation protocol suitable for QA. Future empirical work can test

feasibility, mapping stability, workload implications, and student outcomes across contexts.

References

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code/

American Psychological Association. (2023). *APA guidelines for the undergraduate psychology major: Version 3.0*. https://www.apa.org/about/policy/undergraduate-psychology-major

Bećirović, S., Polz, E., & Tinkel, I. (2025). Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learning Environments, 12*, 29. https://doi.org/10.1186/s40561-025-00384-3

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610-623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer. https://doi.org/10.1007/978-3-319-60648-4

Bruff, D. (2019). *Intentional tech: Principles to guide the use of educational technology in college teaching*. West Virginia University Press.

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220-239. https://doi.org/10.1002/bdm.2155

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281. https://doi.org/10.3102/00346543065003245

Costa, C., Husain-Habib, N., & Reiter, A. (2024). Integrating AI into education: Successful strategies, ideas, and tools from psychology instructors. *Teaching of Psychology, 52*(3), 330-338. https://doi.org/10.1177/00986283241297635

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*(5), 637-643. https://doi.org/10.1007/s12599-019-00595-2

Dratsch, T., Chen, X., Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baessler, B., Sauer, S., Maintz, D., & Santos, D. (2023). Automation bias in mammography: The impact of Artificial Intelligence BI-RADS suggestions on reader performance. *Radiology, 307*(4), e222176. https://doi.org/10.1148/radiol.222176

European Commission. (2025, December 10). *What is ESCO? European Skills, Competences, Qualifications and Occupations (ESCO)*. https://esco.ec.europa.eu/en/about-esco/what-esco

European Federation of Psychologists' Associations AISBL. (2025, July). *EuroPsy: European standard and certificate in psychology*. Regulations. https://www.inpa-europsy.it/wp-content/uploads/2025/11/EuroPsy-Regulations-2025_1.pdf

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1-88. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication No. 2020-1. SSRN. http://dx.doi.org/10.2139/ssrn.3518482

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Kulasegaram, K., Mylopoulos, M., Tonin, P., Bernstein, S., Bryden, P., Law, M., Lazor, J., Pittini, R., Sockalingam, S., Tait, G., & Houston, P. (2018). *The alignment imperative in curriculum renewal. Medical Teacher, 40*, 443-448. https://doi.org/10.1080/0142159X.2018.1435858.

LinkedIn Corporation. (2023, November). *Future of work report: AI at work* (LinkedIn Economic Graph Research Institute White Paper). LinkedIn. https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/future-of-work-report-ai-november-2023.pdf

Mah, D. K., & Groß, N. (2024). Artificial intelligence in higher education: exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education, 21*, 58. https://doi.org/10.1186/s41239-024-00490-1

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review, 35*(113). https://doi.org/10.1007/s10648-023-09823-4

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Rose, D. H., & Meyer, A. (2007). Teaching every student in the digital age: Universal design for learning. *Educational Technology Research and Development, 55*(5), 521-525. https://doi.org/10.1007/s11423-007-9056-3

Sullivan, M., McAuley, M., Degiorgio, D., & McLaughlan, P. (2024). Improving students' generative AI literacy: A single workshop can improve confidence and understanding. *Journal of Applied Learning and Teaching, 7*(2). https://doi.org/10.37074/jalt.2024.7.2.7

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The ethics of algorithms: Key problems and solutions. *AI & Society, 37*, 215-230. https://doi.org/10.1007/s00146-021-01154-8

UNESCO. (2023). *Guidance for generative AI in education and research.* https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

U.S. Department of Labor, Employment and Training Administration. (2025, December 16). *ONET OnLine*. https://www.onetonline.org/

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review, 22*, 271-296. https://doi.org/10.1007/s10648-010-9127-6

White, J., Ng, A., & Maggioncalda, J. (2024). *Generative AI for university leaders* [Online course]. Vanderbilt University & Coursera. https://www.coursera.org/learn/gen-ai-for-university-leaders

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Association for Supervision and Curriculum Development (ASCD).

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Xafis, V., Schaefer, G. O., Labude, M. K., Brassington, I., Ballantyne, A., Lim, H. Y., Lipworth, W., Lysaght, T., Stewart, C., Sun, S., Laurie, G. T., & Tai, E. S. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review, 11*(3), 227-254. https://doi.org/10.1007/s41649-019-00099-x

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

Supplementary Material

Box 1. Illustrative Output for Prompt D: CLO Gap Finder (Course-Level Example)

CLO (verbatim)	**Bloom level**	**KSA tag(s)**	**Assessability note**
Explain core neural mechanisms underlying sensation and perception.	Understand	Disciplinary knowledge	Measurable
Interpret basic findings (figures/tables) from a primary neuroscience article.	Apply	Research literacy; data interpretation	Measurable
Evaluate the strength of evidence in popular neuroscience claims using primary sources.	Evaluate	Critical thinking; research literacy	Measurable
Identify ethical risks and bias implications in AI-generated explanations of neural function.	Analyse	AI ethics; bias awareness	Needs rubric criteria
Use a checklist to verify accuracy of AI-assisted summaries of research articles.	Apply	Human–AI interaction; verification	Measurable
Communicate evidence-based critiques in clear academic writing.	Create	Communication	Measurable

Note. This output is illustrative and provided to demonstrate format rather than report empirical results.

Top gaps (illustrative). Ethics/bias reasoning is present but lacks explicit criteria; verification is not clearly linked to a summative assessment; collaboration/teamwork is not specified (if prioritised at programme level).

Quick fixes (illustrative). Add rubric criteria for ethics/bias reasoning and evidence use; embed verification checklist as a graded component; add an optional group-based critique task aligned to communication/teamwork.

Box 2. Illustrative Output for Prompt K: AI-Supported and AI-restricted Task Pair (With Mini-Rubric)

Target CLO (illustrative). Evaluate the reliability of neuroscience claims using primary sources and appropriate reasoning.

AI-supported task (summary). Students generate a 200–250 word AI-assisted summary of an instructor-provided article, then verify and annotate the summary using a checklist (e.g., flag two unsupported or overgeneralised claims with page/figure evidence). Students submit (a) AI summary, (b) verification checklist, (c) 300–400 word critique grounded in the primary source, and (d) a brief disclosure (tool used; prompt(s) used; what was edited).

AI-restricted task (summary). Students read the same article without AI tools and submit a 300–400 word critique addressing claim, evidence, and limitations (e.g., sampling, measurement, inference), with at least two references to figures/tables/pages.

Comparison/reflection prompt (student-facing). In 150–200 words, compare your AI-supported and AI-restricted critiques: where AI helped, where it misled, which verification steps were effective, and what boundaries you will apply in future AI use to protect accuracy and ethics.

Mini rubric (criteria; 3 levels: Developing / Proficient / Advanced).

1. Evidence use (claims grounded in the primary source).
2. Critical evaluation (limits, uncertainty, inference).
3. Verification quality (AI-supported task only).
4. Ethics/bias reasoning (risks, harms, mitigation/boundaries).
5. Clarity and communication (structure, precision, referencing).
6. Transparency (complete disclosure where applicable).

Table 3. Prompt Playbook Log (Template for Documentation and Quality Assurance)

Field	Entry
Date	
Programme / Course	
Participants (roles)	
Prompt(s) used (A–L)	
Inputs provided (artifacts)	
Tooling context (AI tool / manual-only / local model)	
Key outputs generated (attach/link)	
Accepted outputs (and rationale)	
Disagreements identified (nature, evidence consulted, resolution)	
Rejected/modified outputs (and rationale, including any disciplinary or committee objections)	
Equity and data-governance check	
Decisions made (CLO/assessment/sequencing)	
Actions, owner, due date	
Notes for next review cycle	

Appendix A. Worked Example: Captured Output and Validation Review

This appendix provides a brief worked example based on a captured Prompt D run using a bounded evidence pack for PSY101 Introduction to Psychology, including the module syllabus, five course learning outcomes, a 1,500-word essay brief, and a short marking rubric. The run record documented the prompt, evidence pack, and tool context in line with the Reproducibility and Validation Protocol (the results of this review are detailed in Table 4).

AI Transparency Statement for Appendix A. The illustrative example presented in this appendix was developed using ChatGPT-5 (OpenAI) as a brainstorming and drafting aid. Prompt outputs were reviewed, edited, and verified by the author before inclusion. The example is intended solely to demonstrate the Prompt Playbook workflow and should not be interpreted as a validated curriculum analysis.

Captured Output (Excerpt). The assessment clearly supports application of psychological theory to a real-world issue and provides some opportunity for critical discussion of evidence. However, coverage of research-method strengths, limitations, and ethical implications may be uneven, because these are included in LO2 but are not explicitly required in the essay brief across all topic choices. Breadth of disciplinary coverage is also difficult to verify through a single topic-focused essay. The task assesses written communication effectively but does not provide evidence of oral or collaborative communication. A possible improvement would be to require brief comment on methodological limitations and, where relevant, ethical implications of cited research.

Table 4. Illustrative Validation Review of a Captured Prompt Output

Validation dimension	Observation	Status	Action
Factual integrity	One inference about assessment gaps was plausible but not fully explicit in the evidence pack.	Minor defect	Treated as tentative and checked manually.
Alignment integrity	Most CLO links were plausible, but one cognitive-level judgement was considered slightly too high.	Minor defect	Revised by faculty reviewer.
Completeness/specificity	The recommendation was useful but underspecified in terms of where it should be reflected in the rubric.	Minor defect	Returned for refinement.
Bias/equity review	No obvious harm identified, but any AI-use recommendation would require alignment with local access and policy conditions.	Caution	Reworded to remain tool-neutral.
Decision outcome	Some points were accepted, some revised, and one treated as provisional only.	–	Logged in Decision Log.

This example illustrates the status of Prompt Playbook outputs in the manuscript: they are provisional decision-support artifacts and become curriculum inputs only after documented human validation and revision.

Author Note & AI Use Statement

The author has no actual or perceived conflicts of interest to declare and received no dedicated funding for this manuscript beyond institutional academic time.

This article was conceptually developed and written by the author. During the literature-search phase, AI-enhanced scholarly tools (e.g., Google Scholar with AI features, Semantic Scholar, Consensus) were used to surface candidate sources; all cited works were subsequently selected and reviewed by the author. GenAI tools (including Gemini Pro, Perplexity and ChatGPT-based assistants) were used in an assistive role to generate outlines, alternative phrasings and early drafts of selected sections, tables and prompts, which were then extensively revised and edited by the author. An AI-based reviewing tool (Stanford's Agentic Reviewer) was also used to obtain formative feedback on structure and clarity. AI tools were not used to generate empirical data or to make analytic decisions. The author accepts full responsibility for the content, in accordance with the journal's policy and American Psychological Association (APA) guidance on AI use and citation.