Ravesloot, H., Copeland, A., & Pryor, A. (2026). From chaos to consistency: A human-centered framework for AI-supported assessment at a mission-driven liberal arts college. *Intersection: A Journal at the Intersection of Assessment and Learning, 7*(5), 56–91.
https://doi.org/10.61669/001c.172640

From Chaos to Consistency: A Human-Centered Framework for AI-Supported Assessment at a Mission-Driven Liberal Arts College

Holly Ravesloot, M.S., Amanda Copeland, M.A., Adam Pryor, Ph.D.

Author Note

Holly Ravesloot

Amanda Copeland

Adam Pryor

We have no conflicts of interest to disclose

Intersection: A Journal at the Intersection of Assessment and Learning
Vol 7, Issue 5 Special Issue: The Future of Assessment in an Era of Generative AI

Abstract: This practitioner inquiry reports on a pilot at a small Catholic liberal arts college that used generative artificial intelligence (GenAI)-supported tools to help faculty complete program-assessment documentation during a structured Education Day. Grounded in Catholic social teaching and human-centered AI principles, the pilot tested a framework organized around human judgment, ethical stewardship, and formative human-AI partnership. Survey results from the pilot cohort (*N* = 10; item-level valid responses varied) showed strong faculty endorsement, including unanimous agreement that the process deepened thinking about evidence of student learning and unanimous intent to use AI-supported tools in future cycles. Qualitative findings identified five themes: efficiency and cognitive scaffolding, human-in-the-loop validation, platform reliability and trust, data governance, and institutional support infrastructure. Findings suggest that mission-driven institutions can integrate AI into assessment responsibly when faculty formation, verification checkpoints, and governance structures precede broad implementation.

Keywords: *artificial intelligence, assessment, Catholic higher education, human-centered AI, faculty judgement, program assessment*

Introduction: The Person Before the Algorithm

In Spring 2026, faculty at a small Catholic liberal arts college in the Midwest sat down with a generative artificial intelligence (GenAI) tool to do something most of them had never done consistently: close the assessment loop. What happened next — and how the institution designed the conditions for it to happen — is the subject of this article. In this article, GenAI refers specifically to generative systems used to create or transform content, including the Assessment Wizard and ChatGPT; AI is used when discussing the broader field, institutional governance, or sources that use the broader term.

Pope Leo XIV's Apostolic Letter *Drawing New Maps of Hope,* signed on October 27 and published on October 28, 2025, to mark the 60th anniversary of *Gravissimum educationis* (Second Vatican Council, 1965), opens with a conviction that frames everything that follows: a person "is not a 'skills profile',

cannot be reduced to a predictable algorithm, but is a face, a story, a vocation" (Leo XIV, 2025, §4.1). This is not merely a theological affirmation. It is a design principle — and it stands as the foundational premise of the pilot described in this article.

Program-level learning outcomes assessment has long been one of higher education's most difficult practices to sustain. In a complete cycle, programs articulate outcomes, align courses and assignments, collect evidence, interpret findings, implement improvement actions, and reassess whether those changes strengthened student learning (Banta & Blaich, 2011; Suskie, 2018). At small, resource-constrained institutions, this work is further complicated by high faculty workloads, limited assessment staffing, and competing demands on time. As a result, the gap between collecting evidence and using it to improve learning remains persistent (Banta & Blaich, 2011; Kuh et al., 2015). GenAI is accelerating a disruption that makes this longstanding challenge both more urgent and more tractable. Hathcoat et al. (2026) argue that AI is collapsing the boundaries among learner, test, and assessor, creating what they describe as an "epistemic shift in what counts as evidence and learning" (p. 14). For institutions already struggling to sustain assessment, this shift is not a distant theoretical concern; it is already reshaping how evidence is produced, interpreted, and trusted. The Spring 2026 GenAI Pulse Survey of higher education assessment professionals provides additional context. Among respondents, 79% reported regular or occasional GenAI use, 87% described themselves as self-taught, and 71% reported privacy or ethical concerns; institutional support and faculty collaboration also remained uneven (Slotnick et al., 2026). These results illustrate a field in which individual adoption is advancing faster than coordinated institutional support. Into this moment, Pope Leo XIV's *Drawing New Maps of Hope* (2025) offers a clarifying word for Catholic institutions navigating the intersection of technology and education. "Technologies must serve, not replace, the person," the Pope writes; "they must enrich the learning process, not impoverish relationships and communities" (§9.1). More pointedly, he names AI literacy as one of only three priorities he adds to the Global Compact on Education, calling institutions to "educate in a judicious use of technology and of AI, placing the person before the algorithm and harmonizing technical, emotional, social, spiritual and ecological intelligence" (§10.3). For a Catholic institution designing an AI-supported assessment system, this is not background context. It is a mandate.

The institution chose to answer that mandate deliberately. Grounded in Catholic social teaching and a deep institutional commitment to human dignity, faculty shared governance, and social justice, the institution designed and piloted an AI-supported assessment system that places human judgment, ethical stewardship, and faculty-led interpretation at the center of every phase of the assessment cycle. The result is not a story about AI replacing assessment professionals. It is a story about one institution's decision to let its values, not the technology, drive the design.

This practitioner inquiry was guided by one question: How can a mission-driven, resource-constrained college use GenAI to support program assessment while preserving faculty interpretive authority, ethical stewardship, and formative learning? This article presents the Human-Centered AI Assessment Framework, organized around three principles: human judgment as non-negotiable interpretive authority, ethical stewardship of AI-augmented evidence, and formative human-AI partnership. It then

describes the two interventions through which the framework was enacted: the Assessment Wizard, a customized GenAI-supported conversational tool designed to guide faculty through the institution's program-assessment process, and Education Day, the structured faculty-development setting in which the Wizard was introduced and used. The contribution is both practical and theoretical. Practically, this article offers a replicable model for small, mission-driven institutions seeking to integrate AI into assessment responsibly. Theoretically, it demonstrates that institutional values — in this case, Catholic social teaching — need not constrain AI adoption. They can generate it.

Literature Review and Conceptual Background

The literature informing this study develops across four related areas: the program-assessment cycle and faculty engagement; AI's disruption of evidence and authority; human-centered ethics and governance; and Catholic social teaching as an institutional design framework.

The Persistent Problem of Assessment Practice

Program-level assessment is an iterative, faculty-led process that connects learning expectations to the courses and assignments through which students demonstrate them (Jankowski & Marshall, 2017; Suskie, 2018). In an aligned assessment process, programs begin with what graduates should know and be able to do and then work backward to align program outcomes, course outcomes, learning experiences, assignments, and evidence with those expectations (National Institute for Learning Outcomes Assessment [NILOA], 2018). Curriculum mapping makes these relationships visible and allows faculty to identify gaps, unnecessary repetition, and places where an outcome is taught but not meaningfully assessed (Barthakur et al., 2024; NILOA, 2018). Programs then gather course-embedded evidence, interpret results collaboratively, identify and implement improvements, and reassess whether those changes strengthened student learning (NILOA, 2018; Suskie, 2018).

The value of this cycle lies not in producing reports but in informing curricular and pedagogical decisions. Assessment becomes part of teaching when faculty use evidence to revise courses, sequencing, assignments, pedagogy, curricula, or student supports rather than treating documentation as the endpoint. Decades of research have established the importance of collecting meaningful evidence of student learning, interpreting it thoughtfully, and using it to guide improvement (Banta & Blaich, 2011; Kuh et al., 2015). Yet institutions frequently struggle to move from identifying possible actions to implementing changes and evaluating their effects—a persistent closing-the-loop challenge (Schoepp & Benson, 2016). The problem, therefore, is not a lack of assessment models or evidence supporting their value. It is the difficulty of sustaining complete and meaningful assessment cycles in practice.

Faculty may experience assessment as burdensome, disconnected from teaching, or administratively driven rather than faculty-led (Jankowski & Marshall, 2017). The challenge is therefore developmental as well as procedural: faculty need shared language, time, facilitation, common tools, and opportunities to interpret evidence together (Hoare et al., 2024; Lucander & Christersson, 2020). These challenges can be particularly pronounced at small institutions, where limited assessment staffing, small or single-faculty programs, competing faculty responsibilities, leadership transitions, and

decentralized documentation can disrupt continuity. When adequate supports are absent, participation may become uneven, documentation may vary in quality, and programs may collect evidence without consistently using it to improve student learning.

AI and the Disruption of Assessment

GenAI introduces both opportunities and risks for program-level assessment practice. GenAI tools can help faculty organize assessment evidence, summarize large amounts of information, identify preliminary patterns, and draft documentation more efficiently. For institutions that struggle to sustain complete assessment cycles, these capabilities may reduce the administrative burden that often prevents faculty from moving from evidence collection to interpretation and improvement planning.

Digital assessment scholarship has long argued that technology changes not only how evidence is collected but also the relationships among learners, assessors, tasks, and professional judgments (Bearman et al., 2020). GenAI intensifies this challenge because it can participate directly in the organization, production, and preliminary interpretation of assessment evidence.

Recent scholarship suggests, however, that AI is not simply adding another tool to existing assessment practice. Hathcoat et al. (2026) argue that AI is disrupting foundational assumptions about how evidence is produced, analyzed, and interpreted. In program assessment, this disruption raises questions about how authority is distributed within the assessment process, where accountability resides, and what forms of analysis and documentation remain credible when faculty and AI systems jointly organize evidence and generate preliminary summaries. International guidance similarly emphasizes that responsible AI use in education requires human oversight, transparency, accountability, and protection of personal and institutional data (United Nations Educational, Scientific and Cultural Organizations [UNESCO], 2021). These concerns require institutions to distinguish clearly between AI-supported organization or pattern identification and the professional judgment required to determine what assessment evidence means for a program and its students.

These developments do not replace the established program-assessment cycle. Institutions must still align outcomes, gather appropriate evidence, interpret findings collaboratively, implement improvements, and reassess whether those changes strengthened student learning. GenAI may support selected portions of that work, but responsibility for validating evidence, interpreting results, and determining curricular or pedagogical action remains with faculty and academic leaders. The value of GenAI in assessment therefore depends not only on what the technology can produce, but also on whether faculty are prepared to evaluate and appropriately use its output.

Responsible GenAI-supported assessment requires faculty to possess a baseline level of AI literacy. In this context, AI literacy extends beyond knowing how to enter prompts or generate text. Faculty must understand the capabilities and limitations of generative AI, evaluate outputs for accuracy and unsupported claims, recognize hallucinations, protect student and institutional data, and determine when GenAI-generated summaries require correction, additional context, or rejection (Finney et al.,

2026; UNESCO, 2021). Without this baseline knowledge, human-in-the-loop review may exist as a procedural requirement without functioning meaningfully in practice.

The need for faculty preparation is especially important because AI adoption is occurring unevenly and often ahead of institutional guidance. The GenAI Pulse Survey found that assessment professionals were increasingly using GenAI while relying heavily on self-directed learning and continuing to report concerns about privacy, ethics, and uneven institutional support (Slotnick et al., 2026). Although research on student AI literacy also demonstrates that familiarity with AI tools should not be equated with the ability to evaluate them critically (Finney et al., 2026), institutions cannot assume that faculty are prepared to oversee AI-supported assessment merely because they are willing to use the technology. Structured professional development is needed to establish shared expectations for verification, data protection, transparency, and the appropriate boundaries of AI assistance.

Establishing this baseline does not transfer interpretive authority to the technology. Instead, it equips faculty to exercise that authority more deliberately. Faculty with sufficient AI literacy are better positioned to decide what the tool may assist with, what information must be independently verified, and which interpretations and decisions must remain exclusively human. The challenge is therefore not simply whether AI should be used in assessment, but whether faculty have the knowledge, institutional support, and governance structures necessary to use it responsibly. This need provides the foundation for the Human-Centered AI Assessment Framework developed in the present study.

These needs for human judgment, ethical stewardship, faculty formation, and institutional responsibility are not only technical or professional concerns. At a Catholic institution, they are also mission questions. Catholic educational teaching provides a framework for determining not simply whether AI can be used in assessment, but how its use should serve human dignity, professional responsibility, and the formative purposes of education.

The Magisterial Frame: Drawing New Maps of Hope

Pope Leo XIV's Apostolic Letter *Drawing New Maps of Hope* (2025), published to mark the 60th anniversary of *Gravissimum educationis* (Second Vatican Council, 1965), provides a central magisterial framing for AI in Catholic education. The letter identifies AI literacy as one of only three new priorities Leo XIV adds to the Global Compact on Education, calling institutions to "educate in a judicious use of technology and of AI, placing the person before the algorithm and harmonizing technical, emotional, social, spiritual and ecological intelligence" (§10.3).

The letter's treatment of technology in education is neither technophobic nor uncritical. Leo XIV (2025) affirms that "technological progress is part of God's plan for creation" (§9.2) while insisting that "technologies must serve, not replace, the person; they must enrich the learning process, not impoverish relationships and communities" (§9.1). For Catholic higher education, this is not a new tension — it is the perennial challenge of integral human formation, now applied to a new technological moment.

Critically, the letter speaks directly to the formation of educators. "Technical updates are not enough," Leo XIV (2025) writes; "it is necessary to cultivate a heart that listens, a gaze that encourages, and an intelligence that discerns" (§5.2). This sentence could serve as the mission statement for the institution's Education Day. It names exactly what structured faculty development for AI must cultivate; not just technical competency, but the discernment that makes technical competency trustworthy.

Human-Centered AI and the Ethics of Assessment

The emerging literature on human-centered AI in higher education offers a clear response to this disruption: the answer is not to slow adoption, but to govern it. Luckin (2017) argues that AI in education must augment rather than replace human intelligence, preserving the interpretive and relational dimensions of teaching and assessment that machines cannot replicate. Memarian and Doleck (2024) extend this argument across AI-in-education research, underscoring why human-in-the-loop validation is essential for maintaining epistemic integrity in AI-augmented assessment systems.

Miller (2024), writing in this journal, argues that assessment professionals must adapt to AI not by ceding interpretive authority but by developing new competencies for working alongside AI tools. This is not a passive adaptation. It is an active, professional transformation that requires structured support.

Magnifica Humanitas (Leo XIV, 2026) reinforces this human-centered orientation, insisting that technology “is never neutral,” but always reflects the values and intentions of those who design and deploy it (§9). This claim aligns directly with human-in-the-loop approaches in AI assessment: governance and interpretive authority cannot be external add-ons, but must be embedded within system design.

The Gap This Study Addresses

Despite growing attention to AI in higher education, empirical work on institutionally governed, faculty-led uses of AI in program assessment remains limited. Earlier reviews found relatively little attention to educators’ roles and ethical implications, while more recent synthesis continues to call for stronger ethics, collaboration, and methodological rigor (Bond et al., 2024; Zawacki-Richter et al., 2019). Hathcoat et al. (2026) extend this concern to assessment by calling for a developmental pathway for human-AI partnership. Leo XIV (2025) names a parallel responsibility for Catholic higher education: to offer ‘a diakonia of culture,’ or a service of wisdom, in response to technological change. This study responds to both calls by documenting how one institution translated human-centered and Catholic commitments into a bounded assessment pilot.

Institutional Context and Assessment Problem

About the Institution

The institution is a small Catholic liberal arts college in the Midwest serving a predominantly first-generation, working-class, and majority-minority student population. Its mission emphasizes access, social justice, and the formation of the whole person. Catholic social teaching (CST) is a body of moral

principles that applies the Catholic tradition to social life and institutional responsibility. Three principles were especially relevant to this study: the inherent dignity of every person, subsidiarity, and the preferential option for the poor and vulnerable. In the context of AI-supported assessment, these principles call the institution to avoid reducing students to data points, preserve faculty authority over interpretation and decision-making, and consider how technological choices may affect students, faculty, and programs with the fewest resources.

These commitments shaped the pilot rather than merely providing a theological justification after its design. Human dignity informed the expectation that assessment evidence be interpreted within the context of students' learning and experiences. Subsidiarity supported the use of AI to assist faculty without replacing their professional judgment. The preferential option for the poor and vulnerable reinforced the need for privacy protections, transparent processes, and equitable access to institutional support. Leo XIV's (2025) insistence that a person "cannot be reduced to a predictable algorithm, but is a face, a story, a vocation" (§4.1) captures the central premise of this approach. The institution's mission therefore helped generate the framework's emphasis on human judgment, ethical stewardship, and formative human-AI partnership.

AI Policy and Faculty Preparation Context

At the time of the pilot, the institution had not adopted a comprehensive institution-wide AI policy. Faculty were provided with recommended syllabus language and retained discretion to determine whether, where, and under what conditions students could use AI within individual courses. This approach preserved faculty authority over course design but also meant that expectations and prior experience varied across instructors and programs.

The institution had, however, begun preparing faculty for the changing AI environment before the Spring 2026 pilot. During Fall 2025 Faculty Welcome Week, Lisa Batchelder and Sarah Freye of Michigan State University delivered Call Us Irresponsible: Designing Curricula for the AI Era. The session examined structural opportunities for AI misuse, the vulnerability of traditional assessments, and three possible instructional stances toward AI: intentional non-use, AI as a course tool, and AI as a catalyst for learning. It also encouraged faculty to reconsider learning objectives and assessment strategies in light of the capabilities of GenAI.

On May 6, 2026, Education Day opened with a half-day faculty workshop facilitated by Tawnya Means of Inspire Higher Ed to establish a shared foundation for responsible GenAI engagement before faculty moved into the assessment session.

During the subsequent assessment work, faculty also had access to institutional assessment leadership, technical assistance, and a consultant (author AP). This combination of prior exposure, same-day foundational development, supervised practice, and available consultation was intended to support faculty entering the pilot with differing levels of AI experience. The study did not formally measure participants' prior AI knowledge or use; therefore, these activities should be understood as shared

preparation for the pilot rather than evidence that all faculty began with an equivalent level of AI literacy.

Assessment Infrastructure

The institution had an established program assessment infrastructure before the pilot. Program assessment was an institutional expectation, and academic programs were expected to connect course-embedded evidence to program learning outcomes, interpret the results, identify improvement actions, and document those actions through the institution's assessment cycle. The institution maintained curriculum maps, assessment plans, rubric-based scoring and feedback forms, and end-of-cycle reports aligned with Higher Learning Commission (HLC) Criteria for Accreditation (HLC, 2025). Institutional assessment leadership also provided templates, consultation, and process guidance to support faculty and program leaders.

Despite this infrastructure, participation and documentation quality varied across programs. Some programs completed the required components consistently, while others submitted incomplete documentation or did not clearly demonstrate how findings informed curricular, pedagogical, or student-support decisions. Prior to the pilot, HLC feedback called for clearer documentation of these improvement cycles—particularly more consistent evidence that assessment results were being used to guide action and that the effects of those actions were subsequently evaluated. The central problem, therefore, was not the absence of an assessment process, but uneven continuity, participation, and completion across the full cycle.

Three Challenges That Defined the Problem

Understanding why the institution's assessment practice fell short of its potential requires attention to three related challenges that the pilot was designed to address: continuity, compliance, and consistency.

The first challenge was continuity. The programmatic assessment structure organized academic programs into three reporting cycles, with deadlines distributed across the academic year. Although this model created a predictable calendar, the staggered deadlines sometimes produced confusion and made it difficult to establish a shared institutional rhythm for assessment. Faculty turnover, changes in program leadership, and decentralized documentation further weakened continuity. As a result, evidence gathered during one cycle was not always readily available, interpretable, or usable during subsequent planning and reporting. These conditions pointed to the need for a more unified annual process that connected planning, evidence collection, interpretation, improvement, and reassessment across the full academic year.

The second challenge was compliance, understood here as consistent participation in and completion of the established assessment cycle rather than simple adherence to a reporting requirement. Program assessment was an institutional expectation, and the Director of Assessment provided templates, reminders, consultation, and individualized support to faculty and program leaders. Even with this support, completion varied across programs. Faculty frequently balanced nine-month appointments,

teaching overloads, advising responsibilities, and immediate student needs, while program-leadership turnover, staggered deadlines, and differing levels of familiarity with assessment processes complicated coordination. In some programs, responsibility for gathering evidence, interpreting results, and completing reports was not clearly distributed. Consequently, portions of the assessment cycle were sometimes incomplete, creating gaps in the institutional evidence base and complicating both improvement efforts and accreditation reporting (Jankowski & Marshall, 2017). The challenge therefore reflected the difficulty of sustaining assessment within existing workload, timing, and accountability structures rather than an absence of institutional support.

The third challenge was consistency. Even when assessment activities were completed, the quality, depth, and completeness of the resulting documentation varied across programs. Rubric interpretation and calibration were not always consistent, artifact selection did not always represent the full range of student performance, and findings were not always connected clearly to specific curricular, pedagogical, or student-support actions. Translating evidence into meaningful improvement decisions required sustained faculty interpretation and collaboration—work that competed with the full range of teaching, advising, and service responsibilities. More structured guidance, shared tools, and dedicated time were therefore needed to support greater consistency without replacing faculty judgment.

Together, these challenges were interconnected. Disruptions in continuity made participation more difficult; uneven participation weakened the completeness of the evidence base; and inconsistent processes limited the usefulness of the resulting documentation. The framework described in the next section was developed in response to these conditions.

Human-Centered Framework

Why Framework First

Before describing what the institution built, it is important to explain why it developed a framework rather than simply adopting an available AI tool. The institution's assessment challenges were not primarily technological; they were structural, cultural, and ethical. Introducing AI into a consequential institutional process without an explicit framework could have increased efficiency while leaving unresolved questions about authority, accountability, privacy, and the appropriate role of faculty judgment.

Human-centered approaches to AI emphasize that technology should augment rather than replace professional expertise and that human oversight must remain meaningful throughout AI-supported processes (Luckin, 2017; Memarian & Doleck, 2024). This position also reflects Leo XIV's (2025) warning that educational institutions without a guiding vision risk "soulless efficiency" and the standardization of knowledge (§9.1). The framework therefore came before broad implementation so that institutional values, rather than the capabilities of a particular platform, would determine how AI could participate in assessment work.

The Human-Centered AI Assessment Framework was organized around three principles: human judgment as non-negotiable interpretive authority, ethical stewardship of AI-augmented evidence, and formative human-AI partnership.

Principle 1: Human Judgment as Non-Negotiable Interpretive Authority

The first principle holds that human judgment remains the non-negotiable interpretive authority in program assessment. Assessment is not merely the recognition of patterns in data; it is a process of making meaning from evidence within a particular disciplinary, curricular, and institutional context. Faculty bring knowledge of program outcomes, curriculum, pedagogy, student experiences, and disciplinary expectations to that work. They must therefore determine not only what the evidence shows, but what it means for the program and what actions should follow (Jankowski & Marshall, 2017).

Leo XIV (2025) grounds this principle in the distinctly human dimensions of education, arguing that "no algorithm can substitute what makes education human" (§9.2). Within the framework, AI may help organize evidence, identify preliminary patterns, generate questions, and draft documentation. It may not determine what students have learned, establish final findings, or make curricular and pedagogical decisions. Faculty remain responsible for verifying AI-generated content, adding necessary context, rejecting unsupported claims, and approving the final interpretation.

Hathcoat et al.'s (2026) concept of the "cyborg practitioner"—a professional fluent in both educational assessment and AI literacy—offers useful language for this role. The goal is neither uncritical AI use nor reflexive resistance, but informed collaboration in which technology extends professional capacity while faculty remain firmly in the interpretive position. The extent to which faculty experienced and enacted this principle is examined in the findings.

Principle 2: Ethical Stewardship of AI-Augmented Evidence

The second principle holds that AI-supported assessment must be governed through ethical stewardship. Program assessment may involve student work, institutional records, program-level findings, and other information that requires careful handling. Responsible use therefore requires attention to privacy, data integrity, transparency, and the conditions under which information may be submitted to an AI platform. International guidance similarly emphasizes human oversight, accountability, transparency, and protection of personal information in educational AI use (UNESCO, 2021).

This principle is also consistent with Leo XIV's (2025) call for AI and digital environments to be governed in ways that protect dignity, justice, and participation (§9.3). The framework operationalized these commitments through de-identification requirements, explicit faculty review of AI-generated content, and the expectation that final assessment findings and improvement decisions remain human-led. Faculty were instructed not to submit student-identifiable information and to treat AI-generated summaries as provisional content requiring verification.

Ethical stewardship was therefore understood not simply as an individual faculty responsibility but as an institutional obligation to establish appropriate processes, guidance, and support. This reflects Leo XIV's (2026) emphasis on responsibility, transparency, and governance as necessary conditions for the just use of AI (§44). The findings describe where these safeguards functioned as intended and where the pilot revealed additional governance vulnerabilities.

Principle 3: Formative Human-AI Partnership

The third principle holds that the relationship between faculty and AI should be formative. The purpose of the intervention was not only to produce more complete reports, but also to strengthen faculty members' capacity to engage in assessment, evaluate AI output, and participate in evidence-informed improvement over time.

This principle reflects the Catholic social teaching concept of subsidiarity: technology should support faculty in carrying out their responsibilities rather than assume those responsibilities on their behalf. Leo XIV (2025) similarly argues that technical preparation alone is insufficient and that education must cultivate an intelligence capable of discernment (§5.2). A formative partnership therefore requires faculty to remain active participants—to provide context, question outputs, identify errors, revise language, and determine what should be accepted or rejected.

Meaningful human oversight also depends on faculty having sufficient assessment knowledge and a baseline level of AI literacy. The intended outcome was therefore not automation of assessment, but a supported process through which faculty could engage more confidently and critically in the assessment cycle. Evidence concerning faculty adoption readiness and critical engagement with AI output is presented in the findings and interpreted in the discussion.

Intervention Design: The Assessment Wizard

What the Assessment Wizard Is – And Is Not

Throughout this article, the institutional AI platform refers to the broader technology environment through which the institution provided access to AI-supported tools. The Assessment Wizard refers specifically to the customized conversational assistant developed within that environment to support program-level assessment. A separate program-review assistant was also used for other academic-review tasks. These tools operated within the same broader AI environment but served different purposes. This distinction is important because the platform difficulty described later in the findings occurred in the program-review assistant, not in the Assessment Wizard.

The Assessment Wizard is a GenAI-supported process scaffold designed to assist faculty in completing program-level learning outcomes assessment in alignment with HLC Criteria for Accreditation (CRRT.B.10.010). It is not a grading tool, does not evaluate individual student work, and is not a student-facing application or autonomous decision-making system.

The Wizard was designed as a structured conversational guide that helps faculty move through evidence aggregation, documentation, interpretation, and improvement planning. Its design documentation describes it as "an intelligent partner designed to pull high-quality, actionable insights out of the faculty member through an interview-based approach" (institutional Assessment Wizard Bot Instructions, v3). Its function is to reduce procedural friction, identify missing components, and support reflection while preserving faculty authority over the meaning and use of assessment evidence.

This design reflects the principle that technology should "serve, not replace, the person" (Leo XIV, 2025, §9.1). The Wizard's explicit limitations were therefore not secondary safeguards; they were central to its intended role as a faculty support rather than a substitute for professional judgment.

How the Wizard Works

A faculty member beginning an assessment cycle with the Wizard encounters a guided conversation rather than a conventional reporting form. The Wizard first orients the faculty member to the process, confirms the program outcomes under review, and identifies what evidence has already been collected. Its knowledge base includes institutional materials such as official program outcomes, curriculum maps, and assessment schemas, reducing the need for faculty to re-enter information already available.

The Wizard then guides the faculty member through evidence submission and reflection using a one-question-at-a-time approach intended to reduce cognitive load (Sweller et al., 2019). Its prompts ask faculty to describe the evidence, identify patterns they observe, explain contextual factors that may have influenced results, and consider possible improvement actions. Its tone is designed to be supportive rather than evaluative, recognizing faculty effort while encouraging deeper reflection (institutional Assessment Wizard Bot Instructions, v3).

After evidence has been entered, the Wizard may generate preliminary summaries or identify possible patterns. These outputs are explicitly framed as observations requiring faculty verification and interpretation rather than as final findings. Faculty are expected to compare generated content with source materials, correct inaccuracies, supply disciplinary or institutional context, and determine what should appear in the completed assessment record. The findings section examines how faculty experienced this conversational approach, including its perceived efficiency, collaborative qualities, and limitations.

What the Wizard May Not Do

The Assessment Wizard's design establishes explicit boundaries. It may not generate final assessment findings without faculty review and approval. It may not process student-identifiable information. It may not independently draw conclusions about student learning, determine improvement actions, or override faculty judgment at any point in the assessment cycle.

These restrictions are not merely safeguards placed around the tool; they are central design features of the intervention. Clear limitations and expectations for human validation are important to

trustworthy use of generative AI in educational contexts (Perkins, 2023). A tool that could independently interpret student learning or determine institutional action would no longer function as a scaffold. The Wizard was therefore designed to support the procedural and organizational dimensions of assessment while leaving interpretive authority and responsibility with faculty. Whether these boundaries functioned as intended, and where platform behavior revealed additional vulnerabilities, is examined in the findings and discussion.

Education Day: Formation Before Implementation

Rationale

The Assessment Wizard could not succeed unless faculty were prepared to use it thoughtfully and critically. Faculty development for AI-supported program assessment therefore needed to address both the assessment process and the technology used to support it. Program-level quality work is more likely to become meaningful and sustainable when faculty receive structured guidance, common resources, facilitation, protected time, and opportunities to interpret evidence with colleagues (Hoare et al., 2024; Lucander & Christersson, 2020). AI-focused development adds a second layer: faculty must practice critical evaluation, verification, and ethical use rather than learning only how to generate output (Awadallah Alkouk & Khlaif, 2024). Education Day was designed to bring these two forms of development together.

As described in the institutional context, faculty had received earlier exposure to AI-related curriculum and assessment questions, and the opening workshop on Education Day established a shared foundation for responsible GenAI engagement. The assessment portion of the day moved from that general preparation into applied program-level work. Faculty received their previously completed assessment plans in advance and were instructed to bring de-identified student evidence associated with the program outcomes under review. This preparation allowed participants to work with authentic program materials rather than hypothetical examples.

Education Day was designed as a formation and professional-development experience that included technology training rather than as a conventional software demonstration. Faculty were expected to bring their disciplinary expertise, knowledge of students, assessment judgment, and willingness to question GenAI-generated output to the process.

Design and Implementation

The day began with the half-day faculty workshop facilitated by Tawnya Means of Inspire Higher Ed, which established a common language for responsible GenAI engagement. Faculty considered the distinction between using GenAI as an “answering machine” and engaging with it as a partner, practiced the CRAFT prompting framework—Context, Role, Action, Format, and Tone—and examined the importance of challenging, refining, and verifying GenAI-generated responses.

The program-assessment portion of Education Day then connected that AI foundation to the institution’s Human-Centered AI Assessment Framework. Faculty were introduced to the framework’s

three principles: human judgment as non-negotiable interpretive authority, ethical stewardship of AI-augmented evidence, and formative human-AI partnership. The discussion emphasized how these principles applied to the interpretation of student-learning evidence and to faculty responsibility for final assessment conclusions.
A demonstration of the Assessment Wizard followed, with explicit attention to both its capabilities and its limitations. Faculty were shown how the Wizard could organize evidence, prompt reflection, identify missing components, and generate preliminary summaries. They were also reminded that GenAI tools may produce unsupported claims, misread documents, or generate inaccurate institutional details. Faculty were therefore instructed to approach the Wizard as a critical collaborator rather than an authoritative source.

The majority of the assessment session was devoted to supervised practice. Faculty worked individually and in small groups on their own program assessment reports using the Wizard and the de-identified evidence they had brought. Institutional assessment leadership, technical assistance, and a consultant (author AP) were available to answer questions and support participants as they worked. This combination of protected time, authentic program materials, structured guidance, and immediate consultation was intended to reduce procedural burden while keeping interpretation and decision-making in faculty hands.

By the end of the session, many participants had completed or substantially advanced their assessment documentation. More importantly, the intervention created a shared setting in which faculty could practice organizing evidence, evaluating GenAI-generated content, discussing findings, and connecting results to possible improvement actions. The findings section examines how participants experienced this process and where the implementation revealed both strengths and vulnerabilities.

Methods

Research Design

This study employed a bounded case-study design using a convergent mixed-methods approach. Quantitative survey responses and qualitative open-ended faculty reflections were collected immediately following Education Day and integrated during the interpretation of findings. Institutional process artifacts—including Assessment Wizard design documentation, Education Day materials, implementation notes, and documentation of institutional responses to issues that emerged during the pilot—were used to contextualize the case and trace changes made following implementation. These artifacts were not treated as a separately coded dataset of faculty interactions with AI.

Triangulation across closed-ended survey responses, open-ended reflections, and institutional process artifacts strengthened the credibility of the analysis and reduced reliance on any single source of evidence (Creswell & Plano Clark, 2017; Denzin, 2012).

Case Selection and Boundaries

The case was selected purposively as an information-rich example of a small, mission-driven Catholic liberal arts college piloting an AI-supported process in response to a documented institutional need for greater continuity, participation, consistency, and closure within its program-assessment cycle.

The bounded case consisted of the Spring 2026 Education Day implementation of the Assessment Wizard and the faculty responses collected immediately following the pilot. The unit of analysis was the institutional pilot rather than individual faculty performance or student learning outcomes. Faculty were included in the survey cohort if they participated in Education Day and used the AI-supported assessment process. Faculty who did not participate in the pilot were not included.

Faculty used de-identified student evidence to complete authentic program-assessment work during Education Day. Those student artifacts were not collected, retained, or analyzed as research data. The study therefore examined faculty experiences with the process and the institution's implementation of the intervention rather than the content of individual student work.

Data Sources and Collection Procedures

The primary data sources were a 16-item post-Education Day survey, open-ended faculty reflections embedded within the survey, and institutional process artifacts used to document the intervention and subsequent institutional response. The survey was administered through Microsoft Forms immediately following the pilot to the ten faculty members who participated in the AI-supported assessment process. The complete survey instrument is reproduced in Appendix A.

Institutional process artifacts included the Assessment Wizard instructions, faculty-development and Education Day materials, implementation notes, and documentation of institutional responses to issues described later in the findings. These materials were used to clarify the design and implementation of the intervention and to establish the sequence of institutional actions. They were not analyzed as a separate participant-level dataset.

Ethics, Consent, and Confidentiality

The institution's Institutional Review Board (IRB) has reviewed the submitted proposal for the above-referenced project and determined that the study qualifies for Exempt status. This determination is based on the use of voluntary adult participants, anonymous survey procedures, minimal-risk involvement, and the absence of personally identifiable or sensitive data collection.

Participants were informed through the survey opening statement that their responses would be used for program improvement and could inform a practitioner-research manuscript. Participation in the survey was voluntary. Survey findings are reported descriptively and in aggregate where possible, and qualitative responses are identified only through respondent labels R1–R10.

Student names and other student-identifiable information were not collected for the manuscript analysis. Faculty were instructed to use de-identified evidence during the assessment process, and the

study reports institutional implementation processes and faculty experiences rather than individual student performance.

Positionality

The lead researcher served as both the designer of the pilot intervention and the primary analyst of the resulting data. This dual role provided detailed knowledge of the institutional context and implementation process, but it also introduced the possibility of bias in instrument design, interpretation, and reporting (Herr & Anderson, 2015).

To support reflexivity and provide critical distance, the analysis was reviewed collaboratively by the Director of Assessment and an external consultant. Their perspectives were used to question initial interpretations, examine evidence that complicated the framework's assumptions, and reduce the likelihood that the findings reflected only the lead researcher's expectations (Merriam & Tisdell, 2016).

Survey Instrument

The 16-item AI-Enhanced Learning Evidence and Assessment Cycle Pilot Survey (Table 1) was administered through Microsoft Forms immediately following Education Day. The pilot cohort consisted of ten faculty participants, although item-level response counts varied because some participants did not respond to every question. Results are therefore reported using the valid number of responses for each item.

The survey measured perceived documentation efficiency, interpretive authority, AI assessment literacy, mission alignment, transparency, and future-use intent. Open-ended questions invited participants to describe specific benefits of the AI-supported process, changes they made to AI-generated output, ethical and privacy concerns, desired institutional supports, and their overall experience (Braun & Clarke, 2006; Merriam & Tisdell, 2016).

Item 5 (see Table 1) was designed as a behavioral indicator rather than solely as a perception measure. Instead of asking only whether faculty believed they retained authority, the item asked whether they edited, corrected, added context to, or rejected AI-generated output before using it. This distinction was important because perceived control does not necessarily demonstrate that participants actively exercised human-in-the-loop validation (Memarian & Doleck, 2024).

Survey Item-to-Construct Mapping

Table 1 presents the relationship between each survey item, its response format, and the framework principle or construct it was intended to examine. The complete wording and response options are provided in Appendix A.

Quantitative Analysis

Closed-ended survey responses were analyzed using descriptive statistics. For the five-point items, valid response counts, means, medians, and standard deviations were calculated. Item 5 responses were summarized using frequencies and percentages.

Table 1

Survey Item-to-Construct Mapping

Item	Survey item summary and response format	Primary framework construct
Q1	AI tools helped me organize assessment evidence more efficiently—agreement scale, 1–5	Formative human-AI partnership
Q2	AI tools helped me complete assessment documentation more thoroughly—agreement scale, 1–5	Formative human-AI partnership
Q3	I retained control over interpreting my assessment findings—agreement scale, 1–5	Human judgment
Q4	Describe one specific way the AI-supported tools helped your assessment work—open-ended	Formative human-AI partnership
Q5	I edited, corrected, added context to, or rejected AI-generated output before using it—Yes/No/Sometimes	Human judgment; behavioral indicator
Q6	Describe what you changed in the AI-generated output and why—open-ended (optional)	Human judgment; behavioral indicator
Q7	My professional judgment, rather than the AI, was responsible for final assessment decisions—agreement scale, 1–5	Human judgment
Q8	The process increased my confidence in completing assessment documentation—agreement scale, 1–5	Formative human-AI partnership
Q9	The AI-supported process helped me think more carefully about evidence of student learning—agreement scale, 1–5	Formative human-AI partnership
Q10	The process helped me develop clearer shared language for discussing assessment with colleagues—agreement scale, 1–5	Formative human-AI partnership
Q11	Describe questions or concerns about ethics, privacy, student dignity, or governance—open-ended	Ethical stewardship
Q12	The process respected faculty professional judgment and student dignity—agreement scale, 1–5	Ethical stewardship and human judgment
Q13	The process was transparent about how AI was used and what it could and could not do—agreement scale, 1–5	Ethical stewardship
Q14	What would make you more comfortable or confident using AI in future assessment cycles?—open-ended	All three framework principles
Q15	How likely are you to use AI-supported tools in future assessment work?—likelihood scale, 1–5	Formative human-AI partnership
Q16	Is there anything else you would like to share about your experience with the AI-supported assessment process, the Assessment Wizard, or Education Day?—open-ended	All three framework principles

Note: Agreement scale: 1 = strongly disagree to 5 = strongly agree, Likelihood scale: 1 = not at all likely to 5 = very likely.

Because the pilot cohort consisted of ten participants and valid response counts varied across items, no inferential statistical tests were conducted. Quantitative findings are presented as descriptive evidence of faculty perceptions and reported behavior rather than as estimates that can be generalized to a larger population.

Thematic Analysis

Open-ended survey responses were analyzed using Braun and Clarke's (2006) six-phase thematic-analysis framework: familiarization with the data, generation of initial codes, development of candidate themes, review of themes, definition and naming of themes, and production of the final analysis.

Coding was conducted manually using a spreadsheet and analytic memos. Initial codes were developed inductively from participants' language. Related codes were then grouped into candidate themes and considered deductively in relation to the study's three framework principles: human judgment, ethical stewardship, and formative human-AI partnership.

One participant did not provide open-ended responses; the qualitative analysis therefore draws on nine participant accounts. Credibility was supported through comparison of qualitative responses with closed-ended survey patterns and institutional process artifacts, explicit attention to evidence that complicated the framework's intended outcomes, and collaborative review by all three authors.

Negative Case Analysis and Trustworthiness

Analysis attended explicitly to negative or disconfirming evidence. A participant account involving substantial platform difficulties was retained as a negative case rather than dismissed as an isolated experience. The account was used to examine implementation vulnerabilities, data-governance concerns, and alternative explanations for faculty trust and future-use intent. The details of this experience and the institution's subsequent response are presented in the Findings section so that readers encounter the case after the study design and analytical procedures have been established.

Trustworthiness and potential transferability are supported through the description of the bounded institutional context, the complete survey instrument in Appendix A, the item-to-construct mapping in Table 1, documentation of the timing and conditions of data collection, triangulation across multiple forms of evidence, and explanation of how qualitative themes were developed and reviewed.

Scope of Claims

Findings are interpreted at the level of faculty perceptions, reported behaviors, process quality, and emerging institutional practices. The study does not claim that the intervention caused improvements in student learning or that AI use improved the final quality of program-assessment reports.

Given the small, complete pilot cohort and variation in valid responses across items, no inferential claims are made. The findings are descriptive and context-specific and should be interpreted in relation

to the institution's mission, size, assessment infrastructure, faculty-development history, and implementation conditions (Patton, 2015).

Findings

Quantitative Findings

Table 2 presents descriptive statistics for the closed-ended survey items. Item-level valid response counts varied because one participant did not answer several of the closed-ended questions. Percentage columns were not included because the small and varying valid response counts make the frequencies easier to interpret through the reported means, medians, standard deviations, and valid sample sizes.

Table 2

Descriptive Statistics for Closed-Ended Survey Items

Item	Framework construct	*M*	Mdn	*SD*	Valid *n*
Q1: AI helped organize assessment evidence efficiently	Formative human-AI partnership	4.44	5.00	1.01	9
Q2: AI helped complete documentation more thoroughly	Formative human-AI partnership	4.44	5.00	1.01	9
Q3: Retained control over interpreting findings	Human judgment	4.56	5.00	0.53	9
Q7: Professional judgment—not AI—was responsible for final decisions	Human judgment	4.00	5.00	1.41	9
Q8: Increased confidence in completing documentation	Formative human-AI partnership	4.22	5.00	1.30	9
Q9: Helped participants think more carefully about evidence of student learning	Formative human-AI partnership	4.67	5.00	0.50	9
Q10: Developed clearer shared language for discussing assessment	Formative human-AI partnership	4.22	5.00	0.97	9
Q12: Respected faculty judgment and student dignity	Ethical stewardship and human judgment	4.44	5.00	0.88	9
Q13: Was transparent about how AI tools were used	Ethical stewardship	4.11	5.00	1.45	9
Q15: Likelihood of using AI-supported tools in future assessment work	Formative human-AI partnership	5.00	5.00	0.00	10

Note: See Table 1 for scale information

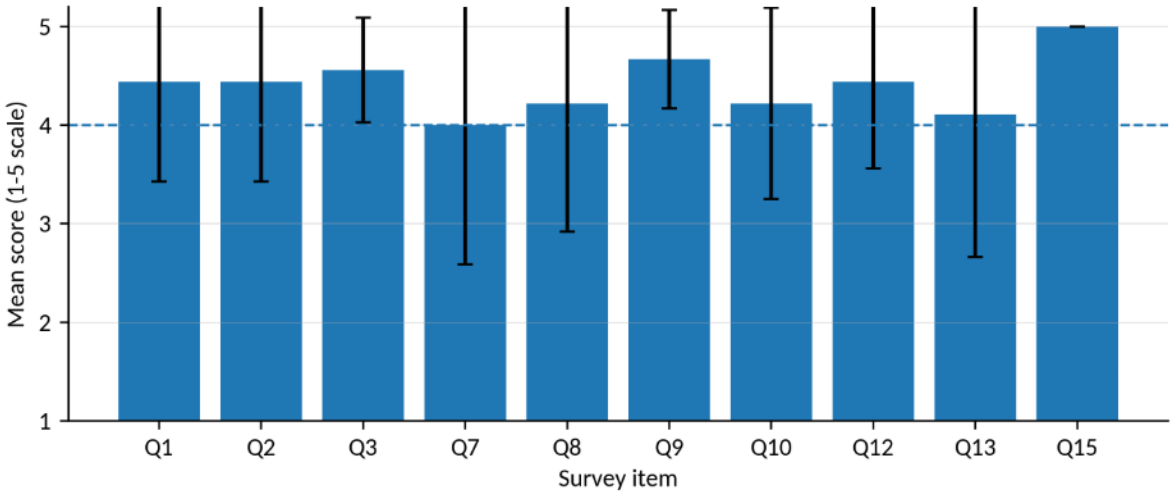
Mean ratings were four or higher for every closed-ended item. The highest mean occurred for future-use intent, with all ten participants selecting the highest response option for Q15 (*M* = 5.00, *SD* = 0.00).

The next highest mean occurred for Q9, which asked whether the process helped participants think more carefully about evidence of student learning (*M* = 4.67, *SD* = 0.50). All nine valid respondents selected either four or five on that item.

Participants also reported retaining control over interpretation (Q3: *M* = 4.56, *SD* = 0.53), with all nine valid respondents selecting either four or five. The greatest variation occurred for Q13, concerning transparency about AI use (*SD* = 1.45), and Q7, concerning whether professional judgment rather than AI was responsible for final decisions (*SD* = 1.41). This variation is considered alongside the qualitative findings concerning platform reliability, trust, and participants' differing experiences with the AI-supported process. Figure 1 visually compares the items' means and their standard deviations.

Figure 1

Mean Likert Scores for the AI-Enhanced Assessment Pilot



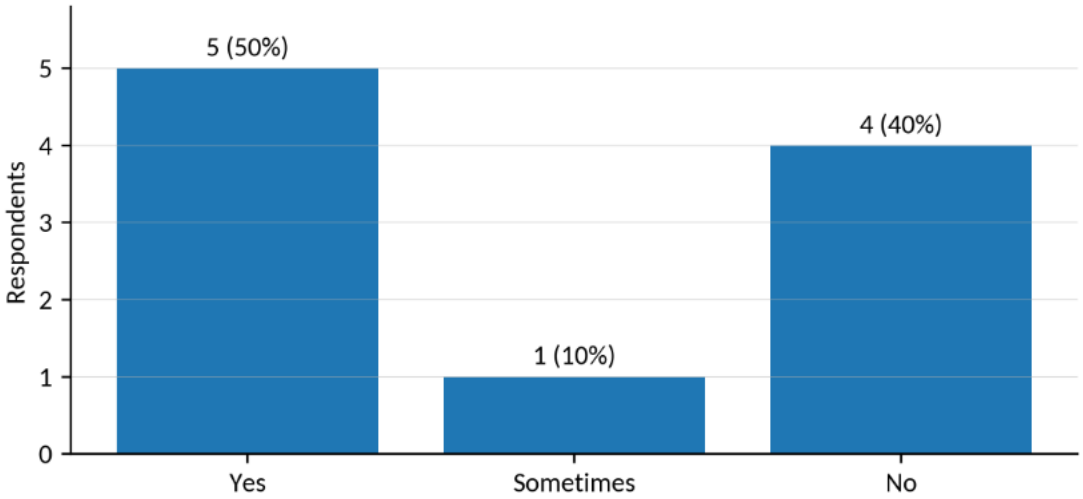
Note: Error bars represent standard deviations; the dashed horizontal reference line marks a mean score of 4.0 on the 1-5 scale.

Figure 2 summarizes responses to Q5, the behavioral indicator asking whether participants edited, corrected, added context to, or rejected AI-generated output before using it.

Six of the 10 participants reported actively validating AI-generated output either consistently or sometimes, while four reported using the output as generated. The survey does not establish whether participants who reported no editing determined that the output required no correction or accepted it without extensive review. Open-ended responses provide additional context concerning the types of revisions faculty made and the risks associated with fluent but inaccurate output.

Figure 2

Human-in-The Loop Validation Behavior



Note. All 10 pilot participants responded to Q5. One participant did not provide qualitative follow-up responses.

Qualitative Findings: Five Themes

Open-ended survey responses were available from nine participants. Thematic analysis identified five themes: efficiency and cognitive scaffolding, human-in-the-loop validation, platform reliability and trust, data governance and ethical stewardship, and institutional support infrastructure. Table 3 summarizes the themes and their principal indicators.

Table 3

Themes Identified in Open-Ended Faculty Responses

Theme	Principal indicators
Efficiency and cognitive scaffolding	Data compilation, alignment summaries, documentation support, speed, and collaborative interaction
Human-in-the-loop validation	Wording revisions, contextual additions, factual corrections, hallucination detection, and rejection of inaccurate output
Platform reliability and trust	Document-processing failures, inaccurate institutional details, platform migration, and differing levels of trust
Data governance and ethical stewardship	Use of institutional information in external tools, uncertainty about data retention, student privacy, and institutional oversight
Institutional support infrastructure	Prompt training, continued practice, technical assistance, knowledge-base maintenance, and escalation procedures

Theme 1: Efficiency and Cognitive Scaffolding

The most consistently reported benefit was the Wizard's ability to organize and compile assessment information, reduce the burden of documentation, and provide faculty with an initial structure for reflection. Participants generally described the tool as responsive and efficient.

R3 reported that the Wizard "compiled the data well" and generated an Excel file. R8 noted that it provided a concise view of how course assignments aligned with learning objectives. R9 described the process as collaborative, explaining, "We just worked together to wrap it up with a bow."

Theme 2: Human-in-the-Loop Validation

Faculty described revisions ranging from stylistic changes to substantive factual corrections. R9 changed wording so that it better reflected the faculty member's voice and corrected an inaccurate course title and an incorrect designation of the course as a capstone.

R10 reported identifying more consequential errors, including fabricated institutional details and inaccurate course classifications. The participant explained, "The dangerous thing is that I was so impressed at first that I was skimming and not reading." This account illustrates how fluent output may initially encourage trust even when the content requires careful verification.

The closed-ended and open-ended findings together show variation in validation behavior. A majority of participants reported editing or sometimes editing AI-generated output, but four reported no editing. The available data do not determine whether this difference reflected the quality of the output, differing levels of scrutiny, or variation in participants' prior AI experience.

Theme 3: Platform Reliability and Trust

After an initially positive Education Day experience, the participant (R10) spent approximately six hours attempting to complete a Faculty Qualifications section using a separate program-review assistant operating within the institutional AI platform. Reported problems included document-processing failures and hallucinated information, such as fabricated degree-granting institutions, invented mission pillars, and inaccurate course categories.

The participant subsequently moved the task to a personal ChatGPT account and reported completing it within approximately two minutes. This experience distinguished confidence in the general usefulness of AI from trust in the institutionally provided platform. R10 also clarified that the Q15 future-use rating of five referred to ChatGPT rather than the institutional AI platform.

Theme 4: Data Governance and Ethical Stewardship

The platform migration also raised concerns about the use of institutional information in external AI systems. R10 reported submitting institutional research materials to a personal ChatGPT account and expressed uncertainty about how the platform might retain or use the information.

Other participants expressed fewer concerns when student-identifiable information was excluded. R3 noted that the absence of student names reduced concerns about ethics, dignity, and privacy.

Together, these responses show that participants' comfort depended partly on the type of information submitted, the platform being used, and the clarity of institutional data-handling expectations.

Theme 5: Institutional Support Infrastructure

Participants identified several forms of support that could strengthen future implementation. R2 requested additional training on effective prompts. R8 emphasized the value of continued use and practice. R9 asked whether updated assessment information would need to be deliberately added to the Wizard or whether the tool would retain information from earlier interactions.

These responses point to two distinct support needs. The first is ongoing faculty development involving prompting, verification, data handling, and responsible AI use. The second is institutional maintenance of the Wizard's knowledge base so that program outcomes, curriculum information, and assessment materials remain accurate across cycles. Participants' responses suggest that continued implementation would require both user support and deliberate curation of institutional content.

Summary: Three Empirically Grounded Claims

Three central findings emerged from the pilot: First, a majority of participants reported actively validating AI-generated output, but validation was not universal. The qualitative responses demonstrate why verification was necessary, particularly when the tool generated inaccurate institutional information.

Second, participants generally reported that the AI-supported process helped organize evidence, support documentation, and encourage attention to evidence of student learning. These findings reflect faculty perceptions immediately following the intervention and do not establish changes in report quality, sustained assessment practice, or student outcomes.

Third, the pilot revealed platform-reliability and data-governance vulnerabilities that were not fully addressed by the initial design expectations. These experiences prompted revisions to the Wizard's instructions and the development of a more structured institutional support process. The implications of these findings for human judgment, faculty formation, and institutional governance are considered in the Discussion.

Discussion

The pilot offers preliminary evidence about how faculty experienced a structured, AI-supported program-assessment process. The findings do not establish that AI improved student learning or the quality of final assessment reports. They do, however, suggest that AI may reduce portions of the procedural burden associated with assessment when its use is paired with faculty preparation, explicit boundaries, human verification, and institutional support. The negative case also demonstrates that

design intentions alone are insufficient: platform reliability, data-governance expectations, and accessible support structures influence whether human-centered principles function as intended.

The Efficiency-Dignity Nexus

Faculty generally perceived the Assessment Wizard as helpful for organizing evidence and completing documentation. Participants reported high agreement that the Wizard helped them organize assessment evidence more efficiently (Q1: *M* = 4.44) and complete documentation more thoroughly than they would have on their own (Q2: *M* = 4.44). These findings are relevant within the institutional context because incomplete assessment cycles were associated partly with limited time, competing responsibilities, and the procedural burden of compiling and documenting evidence.

The findings are consistent with research suggesting that program-level assessment and quality-improvement work are more sustainable when faculty receive structured processes, shared resources, facilitation, protected time, and opportunities for collaborative interpretation (Hoare et al., 2024; Lucander & Christersson, 2020). In this pilot, the Wizard functioned as one component of that broader support structure. It did not replace the assessment cycle or the faculty work of interpreting evidence; it provided a scaffold for organizing information and moving through required documentation.

These reported efficiency gains should not be interpreted as evidence that AI improved the quality of assessment conclusions or deepened faculty reflection. The study did not independently evaluate completed reports or compare them with reports produced without AI. The more defensible interpretation is that the Wizard may have reduced procedural friction and created additional space for faculty to focus on interpretation and improvement planning.

From a human-centered perspective, efficiency and dignity are not necessarily competing priorities. Efficiency has educational value when it supports faculty in exercising professional judgment and engaging more fully with evidence. It becomes problematic when rapid or polished output is treated as a substitute for that judgment. The framework's claim is therefore not that AI improves assessment by itself, but that carefully bounded AI support may reduce administrative burden without displacing the educational purpose of assessment

The Enthusiasm-Vigilance Paradox

The findings reveal an important distinction between retaining authority and actively exercising it. All nine valid respondents agreed or strongly agreed that they retained control over interpreting their assessment findings (Q3: *M* = 4.56). However, six of the 10 participants reported editing, correcting, adding context to, or rejecting AI-generated output either consistently or sometimes, while four reported using the output as generated.

These results do not demonstrate that the four participants who reported no editing accepted inaccurate content. The generated output may not have required revision, or participants may have reviewed it and determined that it was acceptable. The survey did not ask enough follow-up questions to distinguish among these possibilities. Nevertheless, the negative case illustrates why perceived

authority cannot be treated as equivalent to critical verification. Fluent, rapid, and professionally worded output may create an initial impression of reliability even when the content includes unsupported or inaccurate claims.

This tension is described here as the enthusiasm–vigilance paradox: the qualities that make generative AI attractive—speed, fluency, responsiveness, and apparent confidence—may also make users less likely to scrutinize its output carefully. The risk is not simply whether faculty use AI, but whether positive initial experiences lead to passive acceptance of generated content.

The unanimous top rating on the future-use item (Q15: *M* = 5.00) should also be interpreted cautiously. It indicates strong willingness to continue using AI-supported tools, but it does not by itself demonstrate that a formative human-AI partnership was achieved. One participant specifically clarified that the rating referred to ChatGPT rather than the institutional platform. The finding therefore reflects broad adoption readiness toward AI-supported assessment work more clearly than it reflects trust in a particular tool or evidence of sustained critical practice.

Meaningful human-in-the-loop assessment requires more than positioning a faculty member as the nominal final decision-maker. Faculty must possess sufficient assessment knowledge and AI literacy to identify unsupported claims, verify source-based information, recognize hallucinations, protect data, and reject output when necessary (Memarian & Doleck, 2024). Faculty development should therefore include opportunities to practice verification and critical evaluation rather than focusing only on how to generate useful output (Awadallah Alkouk & Khlaif, 2024).

As Slotnick and Boeing (2025) caution, generative AI can support pattern identification and qualitative sense-making while simultaneously introducing risks of fabrication, opacity, and overconfidence. The pilot's hallucination incident reinforces why GenAI-generated interpretations must remain provisional until verified against source evidence.

Future iterations should incorporate facilitated verification checkpoints into the workflow. At designated points, faculty could be asked to compare GenAI-generated summaries with source documents, confirm institutional information, identify unsupported statements, and document substantive revisions before proceeding. These checkpoints would help translate human-in-the-loop design from an abstract principle into a consistent professional practice.

Platform Reliability as a Governance Issue

The negative case represents more than an isolated technical problem, but its location within the institutional AI environment requires clarification. The six-hour difficulty occurred in a separate program-review assistant rather than in the Assessment Wizard itself. The incident therefore should not be interpreted as evidence that the Assessment Wizard produced the more extensive fabricated institutional details reported in that case. The Wizard did, however, generate smaller inaccuracies reported by R9, including an incorrect course title and an incorrect designation of the course as a capstone. Its significance is broader: AI-supported tools operating within the same institutional

environment can create governance risks when expectations for platform reliability, data handling, verification, and technical support are unclear. The available evidence does not establish that legally protected information was disclosed. It does, however, reveal uncertainty about data classification, approved platforms, retention terms, and the appropriate response when an institutionally provided tool does not function reliably. Platform reliability therefore has direct governance implications. When the approved or institutionally supported pathway becomes burdensome, users may migrate toward more familiar tools that appear to work better but operate under different privacy, licensing, and data-use conditions.

This should not be framed primarily as an individual faculty failure. Responsible AI use cannot depend entirely on each user independently interpreting vendor terms, troubleshooting technical failures, or determining which institutional information may be submitted. International guidance emphasizes that human oversight must be supported by transparency, accountability, and clear institutional structures for responsible AI use (UNESCO, 2021). The GenAI Pulse Survey similarly suggests that AI adoption among assessment professionals is advancing more rapidly than consistent institutional support and governance structures (Slotnick et al., 2026).

The institution's submission of a help ticket, revision of the Wizard instructions, and development of a structured support protocol were appropriate responses to the incident. These steps clarified that faculty should pause, verify questionable output, and seek assistance rather than move institutional work independently to an unapproved platform. However, the incident also indicates that a support protocol alone is not sufficient.

A sustainable governance structure should identify which platforms are approved for different categories of information, clarify what data may and may not be uploaded, establish an accessible escalation process, assign responsibility for vendor and platform review, and communicate how the Wizard's institutional knowledge base will be maintained. The safest process must also be the easiest process for faculty to follow. Otherwise, technical friction may continue to push users toward external alternatives.

Mission Alignment as Design Logic

The Human-Centered AI Assessment Framework was developed in conversation with *Drawing New Maps of Hope*. Its human-centered principles were subsequently reinforced by Leo XIV's 2026 encyclical *Magnifica Humanitas,* particularly its emphasis on human dignity, responsibility, transparency, and governance in the use of AI (Leo XIV, 2026).

This alignment should not be presented as empirical validation of the framework. Rather, it clarifies the design logic through which the institution translated Catholic social teaching into operational choices. Human dignity was expressed through the refusal to reduce assessment to automated pattern recognition or treat students solely as data. Subsidiarity was expressed through the expectation that AI support faculty work without assuming faculty responsibility. Ethical stewardship was expressed through de-identification expectations, transparency about AI's limitations, and the requirement that

final interpretation remain human-led. Formation was expressed through the decision to prepare faculty before asking them to use the Wizard.

The principle that technology should "serve, not replace, the person" captures the framework's central boundary (Leo XIV, 2025, §9.1). That boundary also aligns with broader human-centered AI guidance emphasizing augmentation, meaningful human oversight, transparency, and accountability (Luckin, 2017; UNESCO, 2021). The Catholic frame is therefore distinctive without being isolated from the wider ethical literature.

Mission alignment becomes meaningful when it shapes actual decision rights, workflows, data practices, professional development, and institutional support. It is less meaningful when it appears only as aspirational language surrounding the use of a technology. In this pilot, the mission was operationalized through the Wizard's limitations, faculty-led interpretation, preparation before implementation, de-identification expectations, and the institutional response to problems that emerged.

The pilot does not demonstrate that the institution had achieved comprehensive AI governance. It demonstrates a small-scale process through which a mission-driven institution began translating its values into practical design choices. This distinction is important given evidence that assessment professionals often have greater influence within teams than across their institutions as a whole (Slotnick et al., 2026). The pilot moved AI-supported assessment beyond isolated individual experimentation and into a shared institutional process, while also revealing the additional structures required for that process to become sustainable.

Together, the findings suggest that responsible AI-supported assessment depends on the interaction of four conditions: a sound assessment process, faculty capacity to evaluate AI output, clear limits on the technology's role, and institutional structures that support safe and reliable use. The practical implications that follow translate these conditions into specific actions for future implementation.

Practical Implications

Drawing on the five qualitative themes, the quantitative findings, and the three principles of the Human-Centered AI Assessment Framework, Table 4 translates the pilot findings into six practical actions for mission-driven institutions. These recommendations address both the conditions that supported faculty participation and the vulnerabilities exposed during implementation. They are intended to strengthen an existing program-assessment cycle by helping faculty organize evidence, interpret findings, document improvement decisions, and follow those decisions through to reassessment. They should not be understood as a general roadmap for institutional AI adoption.

Together, these recommendations treat AI-supported assessment as an institutional capacity-building project nested within the existing academic assessment cycle. Faculty need protected time, practical guidance, shared tools, structured verification practice, and opportunities to interpret evidence together. The institution, in turn, must maintain clear expectations, dependable support, documented

responsibility for follow-through, and a process for reassessing whether improvement actions were implemented and effective (Hoare et al., 2024; Lucander & Christersson, 2020; Schoepp & Benson, 2016).

Table 4

Practical Implications for Sustaining AI-Supported Program Assessment

#	Practical implication	Evidence and rationale	Recommended institutional action
1	Provide formation before implementation	Faculty entered the pilot with differing levels of AI experience. Prior preparation and the Education Day workshop provided shared language, assessment guidance, supervised practice, and an explicit emphasis on faculty judgment. Program-level quality work is more sustainable when faculty receive structured guidance, common resources, facilitation, and opportunities to work with colleagues (Hoare et al., 2024; Lucander & Christersson, 2020).	Begin each assessment cycle with a structured orientation that connects institutional mission, the program-assessment cycle, AI literacy, data responsibilities, and the limits of the technology. Include guided practice in prompting, verification, and interpretation using authentic assessment materials.
2	Protect time and create a shared assessment workflow	Faculty workloads, staggered deadlines, and competing student-facing responsibilities contributed to incomplete or uneven assessment cycles. Education Day created dedicated time, a common process, and immediate access to assistance.	Establish a unified annual assessment calendar with protected faculty work time, clearly sequenced milestones, common templates, and opportunities for collaborative interpretation. Assign responsibility for each stage of the cycle, including evidence collection, interpretation, improvement planning, and reassessment.
3	Embed structured verification checkpoints	Six of ten participants reported editing or sometimes editing AI-generated output, while four reported using the output as generated. The negative case demonstrated that fluent output could include inaccurate institutional information.	Require faculty to verify AI-generated summaries against source materials at designated points in the workflow. Checkpoints should include confirmation of program outcomes, course information, quantitative results, institutional claims, and proposed improvement actions before content is accepted into the official assessment record.

#	Practical implication	Evidence and rationale	Recommended institutional action
4	Curate institutional knowledge before each cycle	Participants raised questions about whether the Wizard would retain updated information, while the separate program-review assistant incident demonstrated the risks associated with inaccurate institutional information. AI-supported tools therefore depend on current institutional source materials and deliberate knowledge curation.	Assign responsibility for reviewing and updating program outcomes, curriculum maps, assessment plans, institutional terminology, and other source materials before each cycle. Maintain version control and document when information was added, revised, or retired.
5	Establish dependable support and escalation procedures	The platform incident showed that extended technical difficulties may lead faculty to move work to familiar external tools. Participants also requested additional guidance and opportunities for continued practice.	Create a visible support pathway modeled on the institution's LMS help structure. Identify a named contact, publish the help-ticket process, define expected response times, and provide clear instructions about when faculty should stop working, preserve their materials, and seek assistance.
6	Integrate data governance and follow-through into the assessment cycle	The use of internal institutional materials in a personal AI account revealed uncertainty about approved platforms and permissible data. The broader assessment challenge also involved documenting whether improvement actions were implemented and effective.	Provide clear data-classification and de-identification guidance, identify approved platforms, and specify which information may not be submitted to AI tools. At the same time, assign an owner, timeline, and follow-up measure to each improvement action so that the next assessment cycle can determine whether the action occurred and whether it improved student learning.

These actions are consistent with the article's broader claim that mission-driven institutions should approach AI-supported assessment as a practice of discernment rather than merely as a technology implementation. As Leo XIV (2025) writes, Catholic educational institutions are called to become "laboratories of discernment, pedagogical innovation and prophetic witness" (§11.1). In this context, discernment becomes operational through the structures an institution creates: the authority it preserves for faculty, the evidence it requires users to verify, the information it protects, the support it provides, and the responsibility it assigns for acting on assessment findings.

Limitations

This study should be interpreted as an exploratory account of a single, bounded institutional pilot. Although the cohort included only 10 participants, it represented the full group of program directors and department chairs responsible for leading academic programs and program-level assessment at the institution. Collectively, these participants held responsibility across an academic portfolio that included 22 traditional undergraduate degree programs and five graduate degree programs. The cohort therefore provided broad institutional coverage of the academic units involved in program assessment, even though the number of individual participants was small.

Because the study included the complete population of academic leaders participating in the pilot, the findings are meaningful for understanding implementation within this institutional context. The small cohort nevertheless precludes statistical inference and limits transferability to institutions with different missions, governance structures, resources, faculty cultures, assessment infrastructures, or technology environments. One participant did not provide open-ended responses, so the qualitative analysis reflects nine participant accounts.

The lead researcher served as both the designer of the intervention and the primary analyst of the resulting data. This dual role provided detailed knowledge of the institutional context and implementation process but also created the possibility of confirmation bias in the design of the survey, interpretation of responses, and reporting of findings. Collaborative review by the Director of Assessment and an external consultant provided critical distance, but it did not eliminate the influence of the researcher's institutional role.

Data were collected immediately following Education Day. The findings consequently capture participants' initial perceptions, reported behaviors, and intentions rather than sustained use of the Assessment Wizard over time. The study did not examine whether faculty continued to use AI-supported tools in later assessment cycles, whether the intervention improved the quality or completeness of final assessment reports, whether identified improvement actions were implemented, or whether those actions affected student learning. The study also did not measure participants' prior AI knowledge or experience. Differences in familiarity with generative AI may therefore have influenced participants' confidence, validation behavior, platform preferences, and willingness to use AI again.

Several findings were shaped by the broader institutional AI environment. In particular, the most substantial reliability incident occurred in a separate program-review assistant rather than in the Assessment Wizard itself and should not be generalized to the Wizard or to other AI platforms. The reliability concerns reported by one participant may not generalize to other platforms or to future versions of the same system. Similarly, that participant's future-use rating referred to ChatGPT rather than the institutional platform, limiting the extent to which the unanimous future-use result can be interpreted as confidence in the specific intervention technology. Institutional process artifacts were used to document the intervention and contextualize the governance response, but they were not analyzed as an independent dataset of faculty-AI interactions. Claims about faculty experience

therefore rest primarily on survey responses and open-ended reflections rather than systematic analysis of interaction logs, completed assessment reports, or direct observation of faculty behavior.

Accordingly, the study supports cautious claims about feasibility, faculty perceptions, reported validation practices, and emerging governance needs within this institutional context. It does not establish causal effectiveness, durable changes in assessment culture, improvements in student learning, or the superiority of AI-supported assessment over other structured faculty-support approaches.

Conclusion: A Laboratory of Discernment

This bounded pilot documents the feasibility of a human-centered approach to AI-supported program assessment at a small, mission-driven Catholic institution. The findings reflect generally positive initial faculty perceptions, reported validation behaviors, and important governance vulnerabilities rather than evidence that AI improved student learning or the quality of final assessment reports. The case therefore contributes not a demonstration of effectiveness, but an example of how one institution began integrating AI into an established assessment cycle while attempting to preserve faculty judgment, ethical responsibility, and institutional mission.

The pilot also underscores that responsible AI-supported assessment requires more than access to a technological tool. Faculty need sufficient preparation to evaluate GenAI-generated content critically, structured opportunities to verify output against source evidence, clear expectations for the handling of institutional and student information, and dependable institutional support when technology fails. Future assessment cycles should examine whether these structures contribute to more complete documentation, stronger follow-through on improvement actions, sustained faculty engagement, and evidence that implemented changes support student learning. *Drawing New Maps of Hope* informed the Catholic educational context in which this work developed, while Leo XIV's later encyclical *Magnifica Humanitas* further reinforces the framework's human-centered emphasis on dignity, responsibility, and the proper relationship between technological capability and human judgment.

Catholic higher education is called to become, in Leo XIV's (2025) words, a "laboratory of discernment, pedagogical innovation and prophetic witness" (§11.1). This pilot represents one small and provisional step toward that aspiration. For mission-driven institutions, the central question is therefore not simply whether AI can make assessment more efficient, but how technology can be used without surrendering what assessment ultimately requires: human judgment, responsible stewardship, and sustained attention to student learning.

References

Awadallah Alkouk, W., & Khlaif, Z. N. (2024). AI-resistant assessments in higher education: Practical insights from faculty training workshops. *Frontiers in Education, 9,* Article 1499495. https://doi.org/10.3389/feduc.2024.1499495

Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning, 43*(1), 22–27. https://doi.org/10.1080/00091383.2011.538642

Barthakur, A., Kovanović, V., Dawson, S., & Deneen, C. C. (2024). The application of curriculum analytics for improving assessments and quality assurance in higher education. *Australasian Journal of Educational Technology, 40*(4), 20–37. https://doi.org/10.14742/ajet.9383

Bearman, M., Dawson, P., Ajjawi, R., Tai, J., & Boud, D. (Eds.). (2020). *Re-imagining university assessment in a digital world.* Springer. https://doi.org/10.1007/978-3-030-41956-1

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21,* Article 4. https://doi.org/10.1186/s41239-023-00436-z

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research, 6*(2), 80–88. https://doi.org/10.1177/1558689812437186

Finney, S. J., Mehou, J., Miller, S. A., Rotch, R. W., & Miller, J. (2026). Generative AI literacy of incoming American college students: Assessing the need for formal AI literacy training at two institutions. *Research & Practice in Assessment, 21*(3), 22–42. https://www.rpajournal.com/generative-ai-literacy-of-incoming-american-college-students-assessing-the-need-for-formal-ai-literacy-training-at-two-institutions/

Hathcoat, J. D., Slotnick, R. C., & Miller, W. (2026). The end of assessment? Disruption and transformation in the age of AI. *Research & Practice in Assessment, 21*(3), 5–21. https://www.rpajournal.com/the-end-of-assessment-disruption-and-transformation-in-the-age-of-ai/

Herr, K., & Anderson, G. L. (2015). *The action research dissertation: A guide for students and faculty* (2nd ed.). SAGE.

HLC. (2025, Sept. 1). *Revised criteria for accreditation* (CRRT.B.10.010). https://www.hlcommission.org/

Hoare, A., Dishke Hondzel, C., Wagner, S., & Church, S. (2024). A course-based approach to conducting program review. *Discover Education, 3,* Article 5. https://doi.org/10.1007/s44217-023-00085-4

Jankowski, N. A., & Marshall, D. W. (2017). *Degrees that matter: Moving higher education to a learning systems paradigm.* Stylus Publishing.

Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education.* Jossey-Bass.

Leo XIV. (2025, October 27). *Drawing new maps of hope: Apostolic letter on the occasion of the 60th anniversary of the Conciliar Declaration Gravissimum educationis.* The Holy See. https://www.vatican.va/content/leo-xiv/en/apost_letters/documents/20251027-disegnare-nuove-mappe.html

Leo XIV. (2026, May 15). *Magnifica humanitas: On safeguarding the human person in the time of artificial intelligence* [Encyclical letter]. The Holy See. https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html

Lucander, H., & Christersson, C. (2020). Engagement for quality development in higher education: A process for quality assurance of assessment. *Quality in Higher Education, 26*(2), 135–155. https://doi.org/10.1080/13538322.2020.1761008

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1,* Article 0028. https://doi.org/10.1038/s41562-016-0028

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship analysis. *Computers in Human Behavior: Artificial Humans, 2*(1), Article 100053. https://doi.org/10.1016/j.chbah.2024.100053

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Miller, W. (2024). Adapting to AI: Reimagining the role of assessment professionals. *Intersection: A Journal at the Intersection of Assessment and Learning, 5*(4), 99–113. https://doi.org/10.61669/001c.121439

National Institute for Learning Outcomes Assessment. (2018, March). *Mapping learning: A toolkit of resources.* University of Illinois and University of Indiana. https://teaching.temple.edu/sites/teaching/files/images/NILOA%20Mapping%20Learning%20Toolkit-%20updated%20accessible.pdf

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE.

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), Article 7. https://doi.org/10.53761/1.20.02.07

Schoepp, K., & Benson, S. (2016). Meta-assessment: Assessing the learning outcomes assessment program. *Innovative Higher Education, 41*(4), 287–301. https://doi.org/10.1007/s10755-015-9347-2

Second Vatican Council. (1965, October 28). *Gravissimum educationis* [Declaration on Christian education]. The Holy See. https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_decl_19651028_gravissimum-educationis_en.html

Slotnick, R. C., & Boeing, J. Z. (2025). Enhancing qualitative research in higher education assessment through generative AI integration: A path toward meaningful insights and a cautionary tale. *New Directions for Teaching and Learning,* 2025, 97–112. https://doi.org/10.1002/tl.20631

Slotnick, R. C., Boeing, J., & Grillo Pinnelli, B. (2026). *GenAI Pulse Survey for higher education assessment professionals: Spring 2026 results.* Bridgewater State University. https://www.genaiassessment.com

Suskie, L. (2018). *Assessing student learning: A common sense guide* (3rd ed.). Jossey-Bass.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

UNESCO. (2021, November 23). *Recommendation on the ethics of artificial intelligence.* https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International*

Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

About the Authors

Holly Ravesloot, M.S., Director, Online Learning and Institutional Effectiveness, Calumet College of St. Joseph, hravesloot@ccsj.edu

Amanda Copeland, M.A., Instructor, Mathematics, Calumet College of St. Joseph, acopeland@ccsj.edu

Adam Pryor, Ph.D., Founder, Pryor Consulting, adam@pryor-consulting.com

Appendix A

AI-Enhanced Learning Evidence & Assessment Cycle Pilot Survey

Opening Statement: "Thank you for participating in the institution's AI-Enhanced Learning Evidence & Assessment Cycle Pilot. This brief survey (approximately 8–10 minutes) asks about your experience using the Assessment Wizard and AI-supported tools during the assessment process. Your responses will help us understand how AI tools can best support faculty assessment work while preserving your interpretive authority and professional judgment. Responses will be used for program improvement and may inform a practitioner research manuscript. All responses are confidential and will be reported in aggregate or anonymized form."

Section 1: Experience with the AI-Supported Assessment Process

• Q1 (Likert 1–5): The AI-supported tools (Assessment Wizard) helped me organize my assessment evidence more efficiently.

• Q2 (Likert 1–5): Using the AI-supported process helped me complete my assessment documentation more thoroughly than I would have on my own.

• Q3 (Likert 1–5): I felt I retained full control over interpreting my assessment findings throughout the process.

• Q4 (Open-ended): Describe one specific way the AI-supported tools helped your assessment work during this pilot.

Section 2: Human-in-the-Loop Validation

• Q5 (Yes/No/Sometimes): During the process, I edited, corrected, added context to, or rejected AI-generated output before using it.

• Q6 (Open-ended): If you edited, corrected, or added context to the AI output, briefly describe what you changed and why.

• Q7 (Likert 1–5): The AI-supported process made it clear that my professional judgment — not the AI — was responsible for final assessment decisions.

Section 3: AI Assessment Literacy

• Q8 (Likert 1–5): Engaging with the AI-supported assessment process increased my confidence in completing assessment documentation.

• Q9 (Likert 1–5): Working with the AI tools helped me think more carefully about what counts as evidence of student learning.

• Q10 (Likert 1–5): Participating in this process helped me develop a clearer shared language for talking about assessment with colleagues.

• Q11 (Open-ended): What questions or concerns did this process raise for you about using AI tools in assessment? Please share any thoughts about ethics, data privacy, student dignity, or faculty governance.

Section 4: Mission Alignment and Institutional Values

• Q12 (Likert 1–5): I felt the AI-supported assessment process respected faculty professional judgment and student dignity.

• Q13 (Likert 1–5): I felt the process was transparent about how AI tools were being used and what they could and could not do.

• Q14 (Open-ended): What would make you more comfortable or confident using AI tools in future assessment cycles? What support, training, or safeguards would be most helpful?

Section 5: Overall Feedback

• Q15 (Likert 1–5): How likely are you to use AI-supported tools in your assessment work in future semesters?

• Q16 (Open-ended): Is there anything else you would like to share about your experience with the AI-supported assessment process, the Assessment Wizard, or Education Day?

Note on R10, Q15: One respondent clarified that their Q15 rating of 5 pertains to ChatGPT rather than the institutional AI platform: "I put 5 for question 15, but it does not pertain to [the institutional AI platform], it pertains to ChatGPT." This clarification is noted in Table 1 and discussed in Section 8.2, Theme 3.