Examples of Generative AI and Assessment Applications

Dr. Andrew Katz

The Evolving Landscape of Assessment

Traditional Assessment Meets New Realities:

- Balancing robust assessment with student well-being
- Desire for timely, meaningful feedback
- Increasing class sizes and diverse student needs
- Authentic assessment that reflects real-world skills

Focus Today: Practical ways GenAI can *support* and *enhance* (not replace!) our assessment practices.

Why Consider GenAI for Assessment?

- **Opportunities:**

- **Personalized & Scalable Feedback:** Tailor suggestions, provide instant formative input.
- **Diverse & Creative Formats:** Brainstorm novel prompts, scenarios, and assessment types.
- **Efficiency in Design:** Streamline drafting of questions, rubrics, and examples.
- **Student Self-Assessment:** Empower students with tools for revision and practice.

- **Key Considerations (Keep in Mind):**

- Academic Integrity & Authenticity
- Potential for Bias & "Hallucinations"
- Impact on Critical Thinking & Learning

- Equity of Access & Digital Literacy
- **Our Goal:** Leverage the opportunities while navigating the considerations wisely.

---

1. Streamlining Peer Feedback Synthesis

- **The Challenge:** Students often receive a lot of raw, anonymous peer feedback. This can be overwhelming to process and difficult to turn into actionable personal insights for improvement.
- **The "How" (AI-Powered Approach):** Use an LLM to analyze and synthesize anonymous written feedback from multiple peers for an individual student. The AI consolidates common themes, identifies key strengths, and pinpoints areas for improvement.
- **The "Why" (Benefit/Impact):** Transforms raw peer comments into a structured, narrative feedback letter, making the feedback more digestible and less intimidating. The goal is to be more directly useful for student development and reflection.

---

1. Streamlining Peer Feedback Synthesis (cont.)

Basic prompt:

```
"Summarize this feedback: [paste feedback]"
```

Improved prompt:

```
"Summarize the following peer feedback for Student X: [paste feedback].
1. Identify 2-3 key strengths and 2-3 specific, actionable suggestions for improvement.
2. Present as a short letter to the student.
3. Use a supportive, constructive tone.
4. Include specific examples from the feedback to illustrate each point."
```

The improved prompt is better because it:

- Specifies the desired output format (letter)
- Requests a balanced view (strengths and areas for improvement)
- Asks for specific examples

- Sets expectations for tone
- Provides clear structure for the response

---

2. AI as "Critical Friend" for Rubric Refinement

- **The Challenge:** Ensuring a rubric is clear, unambiguous, and consistently interpretable by both students and graders can be difficult. Hidden assumptions or vague criteria can lead to confusion and inconsistent evaluation.
- **The "How" (AI-Powered Approach):** After drafting a rubric, provide it to an LLM. Ask the AI to review the rubric from the perspective of a student or a new grader, specifically looking for ambiguities, unclear language, or areas where criteria might be misinterpreted.
- **The "Why" (Benefit/Impact):** Helps identify potential points of confusion *before* the rubric is used. This leads to clearer expectations for students, more consistent grading, and reduced grading disputes.

---

2. AI as "Critical Friend" for Rubric Refinement (cont.)

Basic prompt:

```
"Check this rubric: [paste rubric]"
```

Improved prompt:

```
"Review this draft rubric for a [Course Name] assignment on [Topic]: [paste rubric].
1. From a student's perspective, what is unclear? Where might they misinterpret criteria?
2. Identify any ambiguous or subjective terms.
3. Suggest 2-3 specific improvements for clarity.
4. For each improvement, explain how it would help students better understand expectations."
```

The improved prompt is better because it:

- Specifies the perspective to take (student's view)
- Focuses on specific types of issues (ambiguity, subjectivity)
- Requests explanations for suggested improvements

- Provides clear structure for the analysis

---

3. "Test Driving" Assignments with AI

- **The Challenge:** Sometimes an assignment prompt, however carefully crafted, might have unforeseen loopholes, ambiguities, or could lead students down unintended paths. It can also be hard to estimate the time or difficulty involved.
- **The "How" (AI-Powered Approach):** Provide your draft assignment prompt to an LLM and ask it to complete the assignment as if it were a student. Analyze the AI's output and its "process" (if you can elicit it through prompting).
- **The "Why" (Benefit/Impact):**
  - Clarity issues in the prompt.
  - Unexpected ways students might interpret the task.
  - Potential for superficial or off-target responses.
  - Whether the scope is appropriate for the allotted time/credit.
  - Helps refine the assignment *before* students receive it.

---

3. "Test Driving" Assignments with AI (cont.)

Basic prompt:

```
"Do this assignment: [paste assignment]"
```

Improved prompt:

```
"Here is an assignment for my [Course] students: [paste assignment].
1. Complete this assignment as if you are a diligent student aiming for a good grade.
2. Show your working/reasoning where possible.
3. Note any points where the instructions are unclear or could be interpreted in multiple wa
4. Estimate how long this would take a typical student to complete.
5. Identify any potential shortcuts or loopholes in the assignment."
```

The improved prompt is better because it:

- Specifies the student perspective to take
- Requests explicit identification of potential issues
- Asks for time estimation
- Looks for assignment vulnerabilities
- Maintains focus on assignment design rather than just completion

---

4. Aligning Assessments with Outcomes & Taxonomies

- **The Challenge:** Ensuring that assessments accurately measure stated learning outcomes and appropriately target desired cognitive levels (e.g., using Bloom's Taxonomy) is crucial for effective course design but can be a complex analytical task.
- **The "How" (AI-Powered Approach):** Provide an LLM with your course learning outcomes, assessment descriptions, and a target taxonomy (like Bloom's). Ask the AI to:

1. Map assessments to relevant learning outcomes.
2. Analyze the cognitive level of assessment tasks according to the taxonomy.
3. Suggest refinements to assessments to better align with outcomes or to target different cognitive skills.

- **The "Why" (Benefit/Impact):** Provides a structured way to review and enhance the alignment of assessments with educational goals. Helps identify gaps or imbalances in how outcomes are assessed and at what cognitive depth.

---

4. Aligning Assessments with Outcomes & Taxonomies (cont.)

Basic prompt:

```
"Check if these assessments match the outcomes: [paste outcomes and assessments]"
```

Improved prompt:

```
"My course learning outcomes are: [list outcomes]. My assessments are: [list assessments].
1. Map each assessment to the primary outcome(s) it addresses.
2. Using Bloom's Taxonomy, what is the primary cognitive level of each assessment?
3. Identify any outcomes that are under-assessed or over-assessed.
4. Suggest one way to modify an assessment to target a higher cognitive level.
5. For any gaps in assessment coverage, propose a new assessment type."
```

The improved prompt is better because it:

- Requests specific mapping between outcomes and assessments
- Incorporates Bloom's Taxonomy for cognitive level analysis
- Identifies assessment coverage gaps
- Suggests concrete improvements
- Maintains focus on alignment and balance

---

5. Generating Exemplar/Model Answers

- **The Challenge:** It's often helpful for students to see examples of what "good" (or "excellent," "adequate," "needs improvement") looks like for an assignment. Creating these varied exemplars, especially for complex tasks, can be time-consuming.
- **The "How" (AI-Powered Approach):** Provide an LLM with an assignment prompt and your rubric (or key criteria). Ask it to generate:
  - An "A" level (exemplary) response.
  - A "C" level (meets basic requirements) response.
  - Optionally, a response that illustrates common pitfalls.
  - Specify desired length or format.
- **The "Why" (Benefit/Impact):**
  - Provides clear benchmarks for students, clarifying expectations.
  - Can be used to calibrate TAs or multiple graders for consistency.
  - Helps students self-assess their own work against concrete examples.
  - Saves faculty time in creating these materials from scratch.

---

5. Exemplar/Model Answers (cont.)

Basic prompt:

```
"Write an example answer for this assignment: [paste assignment]"
```

Improved prompt:

```
"For this essay prompt: [prompt], and this rubric: [rubric criteria for 'A' and 'C'], generat
1. An 'A' level response (approx. 500 words) that demonstrates:
   - Clear thesis and argument structure
   - Strong evidence and analysis
   - Professional writing style
2. A 'C' level response (approx. 500 words) that:
   - Meets basic requirements
   - Has clear areas for improvement
   - Shows common student misconceptions
3. For each response, highlight 2-3 key features that make it 'A' or 'C' level."
```

The improved prompt is better because it:

- Specifies different quality levels
- Provides clear criteria for each level
- Requests explicit identification of key features
- Includes length guidelines
- Focuses on specific aspects of quality

---

6. Developing Complex Scenarios & Case Studies

- **The Challenge:** Crafting rich, authentic scenarios or case studies for problem-based assessments, discussions, or simulations can be very demanding. Ensuring sufficient complexity, relevant details, and varied stakeholder perspectives requires significant creative effort.
- **The "How" (AI-Powered Approach):** Use an LLM as a brainstorming partner or content generator. Provide it with:
  - The core learning objective.
  - Key concepts to be included.

  - Desired context (e.g., industry, historical period, ethical dilemma).
  - Number of stakeholders or variables.
  - Ask it to generate a draft scenario, which you then refine.

- **The "Why" (Benefit/Impact):**
  - Accelerates the creation of engaging assessment materials.
  - Can introduce novel elements or perspectives you might not have considered.
  - Allows for easier generation of multiple, similar-but-distinct scenarios for different student groups or exam versions.

---

6. Developing Complex Scenarios & Case Studies (cont.)

Basic prompt:

```
"Write a case study about [topic]"
```

Improved prompt:

```
"Generate a short (300-word) business ethics case study for an undergraduate course.
1. It should involve a tech company, a new AI product, and conflicting interests between:
   - Profit and corporate responsibility
   - User privacy and public safety
   - Short-term gains and long-term consequences
2. Include 3 distinct stakeholder perspectives with specific concerns.
3. Incorporate 2-3 ethical frameworks that could be applied.
4. End with 3-4 discussion questions that promote critical thinking."
```

The improved prompt is better because it:

- Specifies the context and length
- Identifies key elements to include
- Requests multiple perspectives
- Incorporates theoretical frameworks
- Includes discussion prompts

---

7. Generating Responses for Students to Critique

- **The Challenge:** Students need practice in critically analyzing and identifying flaws in arguments, concepts, and ideas. However, creating examples that contain deliberate, nuanced flaws while maintaining academic rigor can be challenging and time-consuming.
- **The "How" (AI-Powered Approach):**
    1. Use an LLM to generate responses (e.g., arguments, explanations, solutions) that contain specific types of flaws or misconceptions.
    2. Ask students to identify and analyze these flaws, explaining why they are problematic.
    3. Have students suggest improvements or corrections.
    4. Facilitate a class discussion comparing different critiques and approaches to improvement.
- **The "Why" (Benefit/Impact):**
    - Develops students' ability to critically evaluate information and arguments.
    - Provides structured practice in identifying different types of logical fallacies or conceptual errors.
    - Helps students learn to articulate constructive criticism.
    - Creates opportunities for rich classroom discussions about quality and improvement.

---

7. Generating Responses for Students to Critique (cont.)

Basic prompt:

```
"Write something with mistakes about [topic]"
```

Improved prompt:

```
"Generate a 300-word explanation of [concept] that contains 2-3 subtle but significant flaws
1. Be well-written and plausible enough that the flaws aren't immediately obvious to novice s
2. Include one logical fallacy
3. Contain one factual error
4. Have one conceptual misunderstanding
5. Use appropriate academic language and structure
6. Be specific enough to allow for detailed critique"
```

The improved prompt is better because it:

- Specifies the types of flaws to include
- Requests appropriate academic tone
- Ensures the flaws are subtle but identifiable
- Provides structure for the response
- Makes the critique task more focused

---

8. Summarizing Open-Ended Responses for Thematic Analysis

- **The Challenge:** In large classes, analyzing open-ended student responses (e.g., from minute papers, discussion boards, short exam questions) to identify common themes, misconceptions, or areas of understanding can be very time-intensive for instructors.
- **The "How" (AI-Powered Approach):** Provide a batch of anonymized open-ended student responses to an LLM. Ask it to:
  - Identify the 3-5 most common themes or topics mentioned.
  - List any prevalent misconceptions or areas of confusion.
  - Highlight examples of particularly insightful comments.
  - (Caution: Human review is vital for nuance and accuracy).
- **The "Why" (Benefit/Impact):**
  - Quickly provides instructors with a high-level overview of student understanding or sentiment.
  - Helps identify areas where the whole class might need further clarification or discussion.
  - Can inform adjustments to teaching for upcoming sessions.
  - Saves significant instructor time in manually reading and coding all responses.

---

8. Summarizing Open-Ended Responses for Thematic Analysis (cont.)

Basic prompt:

```
"Summarize these student responses: [paste responses]"
```

Improved prompt:

```
"Analyze these [number] student responses to the question '[question text]': [paste anonymize
1. Identify the top 3-5 recurring themes, with:
   - Direct quotes as examples
   - Frequency of each theme
   - Any patterns in how different student groups responded
2. List any common misconceptions, with:
   - Specific examples
   - Potential causes
   - Suggested clarifications
3. Highlight 1-2 particularly insightful responses and explain why they stand out
4. Suggest 2-3 follow-up questions based on the analysis"
```

The improved prompt is better because it:

- Requests specific examples and evidence
- Asks for frequency analysis
- Identifies patterns and misconceptions
- Highlights exceptional responses
- Suggests next steps

---

9. Adapting & Differentiating Assessment Items

- **The Challenge:** Creating varied assessment items that cater to diverse student needs, or generating multiple versions of a quiz/exam to maintain integrity, can be a repetitive and time-consuming task.
- **The "How" (AI-Powered Approach):** Take an existing assessment item (e.g., a multiple-choice question, a short answer prompt, a case vignette). Ask an LLM to:
  - Rephrase it for clarity or simplicity.
  - Change its context (e.g., to a different discipline or real-world scenario).
  - Adjust its complexity level (e.g., for Bloom's: from 'Understand' to 'Apply' or 'Analyze').
  - Generate several plausible distractors for MCQs.
  - Create 2-3 variant questions assessing the same concept.
- **The "Why" (Benefit/Impact):**
  - Facilitates the creation of more tailored and accessible assessments.
  - Supports differentiation for diverse learners.
  - Streamlines the generation of question banks and alternate exam versions.

– Helps refresh existing assessment materials.

---

9. Adapting & Differentiating Assessment Items (cont.)

Basic prompt:

```
"Make this question easier: [paste question]"
```

Improved prompt:

```
"Take this MCQ: [paste MCQ].
1. Rephrase the question stem for greater clarity while maintaining the same cognitive level
2. Generate 2 new plausible distractors that:
   - Reflect common student misconceptions
   - Are similar in length to the correct answer
   - Use similar language patterns
3. Create a similar question that:
   - Applies the same concept to a [different field/context]
   - Maintains the same difficulty level
   - Uses different but equivalent terminology
4. For each modification, explain the pedagogical rationale"
```

The improved prompt is better because it:

- Maintains cognitive level while improving clarity
- Specifies characteristics of good distractors
- Requests explanation of changes
- Ensures equivalent difficulty in variations
- Focuses on pedagogical purpose

---

10. Creating Interactive Practice Modules

- **The Challenge:** Providing students with personalized practice opportunities that adapt to their specific learning needs can be difficult to scale, especially for foundational concepts or skills.
- **The "How" (AI-Powered Approach):** Use an LLM to create the content for a simple interactive tutoring module or a set of adaptive practice questions. For example:
  - Feed it a concept explanation. Ask it to generate a series of questions that check understanding, starting easy and getting harder.
  - Provide common student errors related to a topic. Ask it to generate targeted feedback and follow-up questions for those errors.
  - (Note: True interactivity often requires integration with other platforms, but AI can generate the core content.)
- **The "Why" (Benefit/Impact):**
  - Offers students personalized, self-paced practice tailored to their knowledge gaps.
  - Can reinforce concepts covered in class or provide remediation.
  - Helps students build confidence before high-stakes assessments.

---

10. Creating Interactive Practice Modules (cont.)

Basic prompt:

```
"Write some practice questions about [topic]"
```

Improved prompt:

```
"Create a short interactive quiz on [specific topic].
1. Start with 2 basic recall questions that:
   - Test fundamental concepts
   - Use clear, direct language
   - Have unambiguous answers
2. Then, 2 application questions that:
   - Require synthesis of multiple concepts
   - Present real-world scenarios
   - Have multiple valid approaches
3. For each question, provide:
```

```
   - A hint if the user struggles
   - A detailed explanation of the correct answer
   - Common misconceptions to watch for
4. End with a reflection question that connects to broader course themes"
```

The improved prompt is better because it:

- Specifies different question types and purposes
- Requests progressive difficulty
- Includes support materials
- Addresses common misconceptions
- Connects to broader learning goals

Guiding Principles for Responsible & Effective Integration

As we explore these applications, let's keep some core principles in mind:

- **Transparency & Honesty:**
  - Be clear with students about when and how AI is used in assessment (and when it's not permitted).
  - Discuss attribution and academic integrity openly.
- **Human Oversight & Educator Judgment:**
  - AI tools are assistants, not replacements. Critical evaluation of AI outputs by educators is essential.
  - Final assessment decisions rest with the instructor.

Guiding Principles for Responsible & Effective Integration (cont.)

- **Pedagogical Purpose First:**
  - Focus on how AI can genuinely enhance learning outcomes and support pedagogical goals, not just on using new tech.
- **Equity & Access:**
  - Consider implications for all students. Ensure equitable access to tools if they are required or beneficial.
  - Be mindful of potential biases in AI tools.

- **Start Small & Iterate:**
  - Begin with low-stakes applications to gain experience.
  - Reflect, gather feedback, and refine your approach.

Moving Forward: Your First Step?

Generative AI offers a fascinating, rapidly evolving toolkit for assessment.

The Key: Thoughtful, critical, and pedagogically-driven integration.

A Quick Reflection (for you):

- Which of the examples discussed today resonated most with your teaching context?
- What is **one small, manageable step** you could explore in one of your courses?
  - *(e.g., Try drafting one rubric with AI? Experiment with generating varied question types for a single topic? Test drive one assignment?)*

Questions & Discussion?

Resources & Contact

Stay Connected:

Dr. Andrew Katz akatz4@vt.edu

More Resources:

- IDEEAS Lab: https://ideeas.vt.edu