Measuring Cognitive Growth in AI-supported Learning:

A Literature Review of Educational Assessment Frameworks and Implications for AI-Based Reasoning Systems

Artificial intelligence is rapidly becoming integrated into higher education, not only as a productivity tool but increasingly as a learning companion capable of tutoring, providing feedback, and participating in collaborative reasoning. While recent literature has largely examined the adoption of AI in educational settings, considerably less attention has been devoted to a fundamental question: How should learning itself be measured once AI becomes an active participant in the learning process?

Abstract

This literature review synthesizes research across educational psychology, cognitive science, and learning assessment to examine the principal contracts traditionally used to evaluate development which include: metacognition, conceptual understanding, transfer of learning, self-regulation, reflection, and learning gain (Schraw & Dennison, 1994; Perkins & Salomon, 1992; Rodgers, 2002; Rogaten & Rienties, 2021). Rather than reviewing these contracts independently, this paper examines how each has been operationalized, the instruments used for measurement, and the limitations of current assessment approaches.

The review identifies a common pattern across the literature: most existing instruments rely on self-reporting questionnaires, delayed examinations, reflective writing, or instructor evaluation, all of which provide only indirect evidence of learning. These approaches often capture the outcomes of cognition rather than the reasoning processes through which learning develops.

The paper concludes by proposing that conversational artificial intelligence may provide an opportunity to move beyond static assessment toward continuous observation of reasoning. Rather than replacing established educational measures, artificial intelligence based dialogue systems may complement existing instruments by making previously invisible cognitive processes observable during learning itself.

Introduction

Artificial intelligence continues to reshape nearly every aspect of higher education (Artificial Intelligence in PhD Education; Exploring the Adoption Phenomenon of Artificial Intelligence by Doctoral Students within Doctoral Education). Recent research demonstrates that AI is increasingly being used for literature reviews, programming assistance, writing support, statistical analysis, brainstorming, and individualized tutoring (Artificial Intelligence in PhD Education; Exploring the Adoption Phenomenon...). As it becomes more embedded in our educational environments, researchers have largely focused on questions of adoption, ethics, and academic integrity (Artificial Intelligence in PhD Education). Comparatively little attention has been devoted to understanding how AI may alter the way learning itself is measured.

Educational psychology has for a long time recognized that successful learning reaches beyond memorization, in fact, effective learning involves the development of higher cognitive processes such as metacognition awareness, conceptual understanding, transfer of knowledge, self-regulation, and reflective thinking. These constructs represent different dimensions of cognitive growth that collectively describe how learners acquire, monitor, and apply knowledge across contexts.

Even though each construct has generated extensive research, assessment methods remain surprisingly consistent through each construct. Much like a unifying thread. Most rely upon self-report questionnaires, instructor evaluations, written reflections, or standardized assessments administered after learning has occurred (Craig et al., 2020; Kember et al., 2000; Schraw & Dennison, 1994). These instruments provide valuable insights into learning outcomes but often fail to capture the cognitive processes unfolding while learning is taking place (Craig et al., 2020). The “invisible” side of learning.

With the availability of AI growing exponentially, we must ask the following question: If AI can participate in dialogue with learners, could it also observe, measure, and support cognitive development in real time?

This review examines the educational literature surrounding cognitive assessment to better understand how learning has traditionally been measured and to identify opportunities for AI-supported assessment in future educational environments (National Research Council, 2012). Imagine a classroom in which AI is not only a collaborator, but also a measurement tool helping faculty understand if and how the learning inside and outside the classroom is taking place.

Measuring Metacognition

As we currently understand it, metacognition is the ability to both, be aware of one’s cognitive processes (which is what we define as metacognitive knowledge), and to regulate them (what we understand as metacognitive control). In simpler terms, it is the act of thinking about thinking and monitoring our own cognitive processes (Schraw & Dennison, 1994).

Further defining metacognition, it is important to define meta-knowledge. Meta-knowledge is an upward flow from object-level to meta-level (Fries et al., 1993). Whereas, meta-control is characterized as the downward flow from meta-level to object-level. Therefore, metacognition can be defined as as the top-down control and the bottom-up monitoring of object-level processes

Metacognition has the ability to predict successful learning because learners who actively monitor, evaluate, and regulate their thinking generally perform better academically (Schraw & Dennison, 1994). High performers demonstrate stronger cognitive control and adaptability throughout a multitude of learning activities proven through research.

As of now, metacognition is measured through a couple of tools: The Metacognitive Awareness Inventory (MAI), self-report questionnaires, behavioral tasks, and think-aloud protocols, which are informal classroom Q&A style discussions (Schraw & Dennison, 1994).

There are limitations that present themselves in these forms of measurement we currently use. These methods require a heavy reliance on introspection which lead to limited external validity, they can also be difficult to observe, and therefore measure, in real time, and the current educational methods differ vastly from the neuroscience measures which rely heavily on scanning devices such as MRI imaging (Craig et al., 2020).

This is an area in which we have the opportunity to introduce AI. Instead of asking students to reflect on their experiences after they have completed an assignment we could have AI observe the experience as it is happening. AI could be a part of the monitoring during the dialogue with the student and it could be evaluating what it observes.

Measuring Conceptual Understanding

Humans perceive conceptual understanding as the “knowing of facts”, but that would be a misrepresentation since conceptual understanding is when a learner demonstrates the ability to connect ideas, explain why something is true, transfer this knowledge into new situations, and monitor their own thinking (OpenLearn, n.d.).

Knowing facts is much more akin to memorization which is usually narrower (OpenLearn, n.d.). It can support the learning process, yet on its own it does not produce the correct recall without robust explanation, and in some cases robust application.

We currently use the following methods to measure conceptual understanding: concept inventories, open-ended questions, misconception analysis, and instructor interviews (OpenLearn, n.d.). These methods present several challenges for the accurate measure of conceptual understanding, the first one being that they only capture the final answer, none of them capture the reasoning process.

As aforementioned, this is an area in which AI could be introduced to overcome said challenge. We could introduce AI as a monitoring tool to gain further understanding of the planning and reasoning process in each student that engages with the chatbot. We could also have AI introduce dynamic questioning to probe misconceptions until a genuine understanding is reached by the student, further improving the learning process and possibly the learning outcomes.

Measuring Transfer of Learning

Transfer of learning is one of the defining characteristics of deep learning. It is the process of learning in one context affecting the performance in another (Perkins & Salomon, 1992). It can be further defined as the extent to which knowledge, skills, and abilities are applied, generalized, and maintained over time (National Research Council, 2012).

Transfer of learning can also occur in two different ways: near transfer and far transfer (Barnett & Ceci, 2002). Near transfer is when what has been learned is applied to a very similar situation (Barnett & Ceci, 2002). This requires less knowledge and understanding. Far transfer is when what has been learned is applied to a situation that is meaningfully different, and requires a higher level of knowledge and understanding of the topic learned (Barnett & Ceci, 2002).

We measure transfer learning by encountering new problems in our day to day, by participating in delayed assessments such as self-learning questionnaires administered at the end of a completion of an experience or a task, and through authentic tasks (Perkins & Salomon, 1988).

In the measurement of transfer we face one challenge which is that transfer is only measured after learning has finished. This presents an opportunity in which AI could introduce novel scenarios in real time to determine whether learners apply concepts beyond memorization examples and demonstrate ownership of knowledge.

Measuring Reflection

The final cognitive process that this paper will explore is reflection. We can define reflection as the deliberate process of examining experience to build meaning (Rodgers, 2002). Research has proven that reflection deepen the understanding of concepts and builds self-regulated learners (Boud et al., 1985). We also define reflection as a deliberate self-interrogation of an experience far more complex than a simple summary or a review of what happened (Rodgers, 2002).

It is recommended to engage in frequent brief periods of reflection, two to five minutes, immediately after a task is completed for the practice to be more effective than assignments completed long after the task (Cambridge International, n.d.). It is also a recommendation for reflection to engage emotions since it helps students identify any frustrations or surprises signaled during the process of understanding and learning (Boud et al., 1985).

Journal entries promote reflection and go hand in hand in promoting more effective cognitive learning processes this is because entries give reflection a cumulative structure (Kember et al., 2000). Students, ideally, should be prompted with questions that escalate in depth throughout the semester, or during the duration of the experiment, to help them grow depth of emotional connection and understanding with their cognitive processes. Early entries should be of descriptive nature such as responding to an experience, whereas entries further into the process should evaluate own thinking and consider how the perspective has changed over time and if that was affected by reflection or not.

It is important to point out that, in opposition of common beliefs, reflection is not an add-on to active learning (Boud et al., 1985). Reflection is a mechanism by which active learning produces understanding which can lead to far transfer of learning. Without a reflective component, students will be engaged socially and physically but may not consolidate learning at the conceptual level.

Calling back to the beginning of this literature review, metacognition and reflection happen to be deeply intertwined (Rodgers, 2002; Schraw & Dennison, 1994). Metacognition is the capacity to monitor and regulate our own thinking and reflection is the primary tool through which that capacity can be exercised and further developed. Self-regulated learning depend son reflection at every single state of the regulatory cycle (Schraw & Dennison, 1994). The cycle begins with learners setting goals, monitoring progress, evaluating outcomes, and finally adjusting strategies.

Without reflection, student would face a lack the knowledge of the self necessary to regulate effectively.

Measuring Learning Gain

Researchers usually define a learning gain, or cognitive growth, as a change over a period of time in which knowledge, skills, reasoning, or related competencies experience what we perceive as a positive change (Rogaten & Rienties, 2021).

In higher education "learning gain" is often used as a umbrella term commonly defined as improvement made during a process of learning such as university or college (Rogaten & Rienties, 2021).

Research proves that there are at least three common ways to define growth:

1. Treating it as a "distance traveled" which points to having traversed a difference in learner status at two different points in time.

2. Perceiving a value being added from starting point to prior attainment in performance

3. There is a considerable improvement in knowledge, skills, work readiness, and or personal development.

(Rogaten & Rienties, 2021)

These changes are often measured longitudinally but they can also be measured with pre/post or value added models (Rogaten & Rienties, 2021).

This measurement is performance based in measurement, because researchers frequently rely on assessments, rubrics, or test scores (Rogaten & Rienties, 2021).

Discussion

Across every construct examined in this review, a consistent pattern emerges. Educational researchers have developed sophisticated theories describing cognitive development, yet most assessment instruments remain retrospective (Schraw & Dennison, 1994; Perkins & Salomon, 1992; Rodgers, 2002; Rogaten & Rienties, 2021). They measure what students report, remember, or produce after learning has occurred. Far fewer approaches directly observe cognition as it unfolds.

Recent advances in conversational AI introduce the possibility of a different assessment paradigm. Rather than relying exclusively on questionnaires, exams, or written reflections, AI systems may be capable of engaging learners in structured dialogue that reveals evidence of metacognition, conceptual understanding, transfer, reflection, and self-regulation during the learning process itself (Artificial Intelligence in PhD Education).

This observation does not suggest replacing existing educational assessments. Instead, it suggests expanding them. AI supported dialogue could complement traditional instruments by making aspects of cognition that have historically remained invisible more accessible for study.

Research Gap

This literature review exposes an opportunity to research a better way to implement educational assessment measures across all themes aforementioned surrounding the measurement of cognitive growth. This is a glowing opportunity to apply AI systems that may be capable to measuring the process of thinking itself as it is happening, instead of having to wait for a task, experiment, experience to be completed before a self-questionnaire can be administered to attempt to accurately measure cognitive growth. Current educational assessment measures fall short when it comes to this topic. This is when we introduce AI.

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Boud, D., Keogh, R., & Walker, D. (Eds.). (1985). *Reflection: Turning experience into learning*. Kogan Page.

Cambridge International. (n.d.). Reflective practice and reflection.

Colley, S. M., Tremblay, K. J., & Selig, J. P. (2012). Reflection and critical thinking: A key component of thinking critically. *The Journal of Scholarship of Teaching and Learning, 12*(4), 1–18.

Craig, K., et al. (2020). Evaluating metacognitive self-reports: Systematic reviews of the value of self-report in metacognitive research.

Fries, S., Diehl, M., Kirschbaum, C., & Hell, B. (1993). A meta-cognitive scale: Development and validation.

Kember, D., Leung, D. Y. P., Jones, A., Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Wong, M., & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education, 25*(4), 381–395.

National Research Council. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century and related chapter on teaching and assessing for transfer.

OpenLearn. (n.d.). How can teachers assess students’ conceptual understanding?

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership, 46*(1), 22–32.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education*.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842–866.

Rogaten, J., & Rienties, B. (2021). Learning gain in higher education: Definitions, measurement, and implications.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460–475.

If you want, I can turn this into a fully formatted APA 7 reference list with hanging indents and only the sources that are actually strongest for your topic.