NEPC Review: Productive Struggle: How Artificial Intelligence is Changing Learning, Effort, and Youth Development in Education (Bellwether, June 2025)

Reviewed by:

Thomas M. Philip
University of California, Berkeley

October 2025

National Education Policy Center

School of Education
University of Colorado Boulder
nepc.colorado.edu

Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Mark Paige
Academic Editor

Elaine Duggan
Production Design

Alex Molnar
NEPC Director

Kevin Welner
NEPC Director

Suggested Citation: Philip, T.M. (2025). *NEPC Review: Productive struggle: How artificial intelligence is changing learning, effort, and youth development in education*. Boulder, CO: National Education Policy Center. Retrieved [date] from https://nepc.colorado.edu/review/struggle

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

The National Education Policy Center (NEPC), a university research center housed at the University of Colorado Boulder School of Education, sponsors research, produces policy briefs, and publishes expert third-party reviews of think tank reports. NEPC publications are written in accessible language and are intended for a broad audience that includes academic experts, policymakers, the media, and the general public. Our mission is to provide high-quality information in support of democratic deliberation about education policy. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence and support a multiracial society that is inclusive, kind, and just. Visit us at: http://nepc.colorado.edu

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Productive Struggle: How Artificial Intelligence is Changing Learning, Effort, and Youth Development in Education (Bellwether, June 2025)

Reviewed by:

Thomas M. Philip
University of California, Berkeley

October 2025

Summary

Productive Struggle: How Artificial Intelligence is Changing Learning, Effort, and Youth Development in Education, recently released by Bellwether, considers the role of GenAI in education. It proposes a criterion for evaluating these new technologies' impact on student learning: When does ease afforded by GenAI enable greater learning, and when is ease a shortcut with a hidden cost? Toward this end, it examines how AI-powered tools may interact with a student's memory and information processing, attention and engagement, motivation and mindset, and metacognition and self-regulation. Unfortunately, the report conflates the distinct and often implicit meanings of learning, which leads to unjustified conclusions about the potential impact of AI in education. Taking a highly individualistic lens to learning, and thus often overlooking complex classroom environments that impact learning, the report also makes unfounded inferences about the use of AI-powered tools in classrooms. Further, it does not offer proactive recommendations to mitigate the known harm associated with the use of GenAI, particularly when used with children and youth. In sum, while the report restates and reminds us of important questions that must be addressed, its usefulness for policymakers is quite limited.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Productive Struggle: How Artificial Intelligence is Changing Learning, Effort, and Youth Development in Education (Bellwether, June 2025)

Reviewed by:

Thomas M. Philip
University of California, Berkeley

October 2025

I. Introduction

Access to Generative AI (GenAI) chatbots like ChatGPT has upended schools. Mainstay homework and course assignments that constituted the bulk of grades in middle and secondary schools are now easily completed with chatbots. The difficulty to accurately detect the use of GenAI has generated a crisis, with educators even more anxious about cheating and plagiarism. An added concern is that AI-powered tools can short-circuit vital learning processes, making students dependent on the technology. On the other hand, GenAI could potentially reduce time spent on certain tedious tasks, freeing up resources for higher-order thinking and more meaningful activities.

Thus, GenAI presents profound opportunities, challenges, and dilemmas for schools. But how to balance the potential risks with benefits is worrisome, especially as AI entered schools "without evaluation, assessment of risks and benefits, training for educators, or any other adoption steps that historically would have been considered indispensable to effective technology integration."[1]

Attempting to address this gap is a recent Bellwether report, *Productive Struggle: How Artificial Intelligence is Changing Learning, Effort, and Youth Development in Education*, authored by Amy Chen Kulesa, Marisa Mission, Michelle Croft, and Mary K. Wells. It endeavors to go beyond the popular good/bad binary of GenAI in classrooms. Drawing on the construct of productive struggle—the process in which a learner engages with "challenging tasks or problems that require effort, critical

thinking, and persistence to solve"[2]—it proposes a criterion for evaluating these new technologies: When does ease afforded by GenAI enable greater learning, and when is ease a shortcut with a hidden cost?

II. Findings and Conclusions of the Report

The report finds that the risk with GenAI "is more than cheating; it is about students outsourcing the hard, mental work, like generating ideas or grappling with ambiguity, that builds their capacity to think independently."[3] This perspective is grounded in a framework of productive struggle, which entails students working through problems that are beyond the reach of what they can solve independently but can reasonably succeed at with support. The level of struggle needs to be appropriate: Too little struggle does not carry significant learning benefits and too much can discourage learning.

Drawing on research in cognitive science, the report identifies four factors that contribute to learning[4]: memory and information processing; attention and engagement; motivation and mindset; and metacognition and self-regulation. It describes how productive struggle can contribute to each of these processes and how AI-powered tools can either benefit or deter productive struggle in each. For instance, with respect to metacognition, the report suggests that an AI-powered tool's timely and targeted incorporation of guided self-reflection into a lesson can increase a student's capacity for metacognition. Under other conditions, over-reliance on AI guidance can diminish students' ability to manage their own learning process.

The report identifies seven recommendations, emphasizing the need for educators, developers, funders, and policymakers to work together to intentionally integrate GenAI tools into schools:

1. **Reimagine and redefine what students need to know and become.** Educational goals need to "cultivate students' capacity to make meaning, weigh evidence, sustain effort, and exercise critical discernment and complexity."
2. **Build coherent systems that align capacity and technology to learning.** Effective AI adoption requires "a coordinated approach across professional development, infrastructure, curriculum, and assessment."
3. **Empower educators to redesign assignments for an AI-rich world.** Educators need support in developing new assessments that help students "learn not just content but also discernment, iteration, and independent thinking."
4. **Reinvest in research that reflects the moment.** There is a need for a

"new wave of interdisciplinary inquiry, bringing together cognitive scientists, developmental psychologists, educators, and technologists, to study how students actually experience AI in real classrooms."

5. **Reorient measurement to learning, not just use.** Educational technology should be evaluated on outcomes like engagement, knowledge transfer, differentiation, and classroom connection rather than conventional metrics like number of users or session length.
6. **Develop benchmarks that reflect how students learn.** AI tools should be assessed for "how they shape persistence, curiosity, and long-term understanding across diverse students, including those with learning differences and those learning in multiple languages."
7. **Center learning science in product design.** Principles from cognitive science, developmental psychology, Universal Design for Learning, and the science of motivation and productive struggle should be embedded into the design of tools from the outset.

III. The Report's Rationale for Its Findings and Conclusions

The driving assumption behind the report is that productive struggle is a key process involved in meaningful learning. It focuses on four components of learning that it identifies as most crucial: memory and processing; attention and engagement; motivation and mindset; and metacognition and self-regulation. It assesses AI's possibilities and limitations within each of these domains, although it is not altogether clear how the domain assessments are specifically linked to the final recommendations.

IV. The Report's Use of Research Literature

The report incorporates peer-reviewed research, but heavily relies on reports, blog posts, and news articles. Its use of the research is limited by its narrow definition of productive struggle and learning, which are mostly conceptualized as individual attributes and processes. It relies exclusively on psychological and cognitive science approaches to learning. While these perspectives have value, they often overlook complex classroom environments that impact learning.[5]

For instance, it is assumed that students' zones of productive struggle can be easily identified—an assumption that holds up better in individual experimental conditions rather than social environments like a classroom. Educational researchers who study classrooms have raised concerns about such individualistic views of effort and

motivation.[6] Realistically, productive struggle involves an interplay of peer dynamics, student-teacher interactions, and the particular task at hand.[7] Productive struggle must be viewed within a collective enterprise that depends on a host of social factors: how students engage with each other's ideas, collaboratively determine the parameters of a problem, and negotiate social relations such as authority, smartness, and interpersonal issues.[8]

Additionally, the report creates a false distinction between cognition and "the human side of learning," a dichotomy that has been contested in the literature.[9] It mistakenly extrapolates the potential affordances of individual students' use of AI tools to whole class contexts without a basis for doing so. Such unsubstantiated conjectures gloss over the centrality of expert teacher facilitation.[10] Relatedly, emphasis on individualistic over collaborative approaches to learning reflect different assumptions about and goals for education, especially related to democracy.[11]

The report's lack of attention to two areas of research is particularly glaring. It fails to adequately highlight the associated risks to student privacy, a prominent theme in existing research.[12] It also entirely sidesteps a well-developed line of literature documenting educational technology's negative and disproportionate impact on low-income students of color.[13]

V. Review of the Report's Methods

The report does not outline its methods for selecting documents nor its criteria for what constitutes learning. Specifically, the report lacks a clear definition of learning, which makes it difficult to assess findings across the literature. To be sure, learning has many meanings: the acquisition of skills and knowledge, conceptual change, and changes in identity or participation, to name a few.[14] Studies referenced in the report use implicit definitions, but this only contributes to further confusion, leaving one to wonder what form of learning the report is referencing at any given moment. An attentive reader is left comparing apples to oranges with respect to claims about learning. To the extent we can discern a definition, it appears to default to test scores and self-reports. These are poor measures of deep conceptual learning.

Additionally, the report does not adequately attend to differences in learning over time. A cited study, which reported on a randomized control trial in high school math classrooms, highlights why it is important to consider time. The study showed that ChatGPT can dramatically increase students' immediate success in assignments where they used the tool, but degraded their performance relative to the control group when they did not have access to the tool in an examination.[15] Looking at learning over the short term alone would be misleading. To be clear, it is not simply a matter of long-term retention. Without a systematic consideration learning over

time, the report conflates fundamentally different types of learning.

Without a baseline definition of learning, or even an assessment of how the cases in the cited literature measure up to the report's criterion of productive struggle, it is not possible to compare the claims of each cited case.

VI. Review of the Validity of the Findings and Conclusions

The report's conclusions call for educators, developers, funders, and policymakers to work together to address the intentional adoption of AI into classrooms, develop an interdisciplinary research base for how AI shapes learning opportunities, support teachers, and expand notions of learning and assessment. These are all reasonable and laudable. But they are not based on findings emanating from the report. Put differently, the conclusions (which are rather generic and aspirational) could have been written apart from this report.

As noted above, its most serious flaw is its failure to explicate the forms of learning afforded by an AI-powered tool. Without this, how can we answer the central question asking for an assessment of AI's impact on that very issue? The report applies a concept of productive struggle, which could have proven helpful. But it views the struggle as individual in nature, overlooking classroom dynamics that impact learning. These include, of course, the complex interplay of different students in each setting, peer and student relationships, and collaboration, among others.

The report's effortfully balanced approach prevents it from providing a more in-depth exploration of the very serious risks of GenAI in classrooms. It could have maintained its fairness without glancing over topics where AI poses great potential harm, still underscoring the documented risks given the stakes. For instance, the report extends into areas "beyond cognition," what it terms "the human side of learning." Yet on the "human side" is the real, documented role of chatbots in suicidal ideation. Unfortunately, the report does not give this threat sufficient weight, writing that "while one study found that chatbots can reduce suicidal ideation in some users, other cases, highlighted in recent lawsuits, point to potential harm for other users, particularly those under age 18."[16]

This risk to student life raises the stakes of any study quite beyond whether AI-powered tools may be "beneficial" or "harmful." The risks of students using chatbots are numerous, and corporate developers have often not been trustworthy. As one example among many, Meta formulated policies for chatbot behavior that permitted the company's tools to "engage a child in conversations that are romantic or sensual."[17] Given all of this, there must be guardrails that protect children and youth in any recommendations around AI use, even if it slows technology development.

AI-powered tools' potential harm is not limited to self-injury. They can also harm youth through a multitude of routine and cumulative interactions: from having to use image generation tools that exhibit racial biases[18] to learning with tools that covertly discriminate based on racialized dialects.[19]

In total, the existence of these quite substantial risks makes the report's even-handed approach hard to swallow. From learning to suicidal ideation, the report notably lacks any recommendations that ensure the mitigation of harm from the use of GenAI.

VII. Usefulness of the Report for Guidance of Policy and Practice

The report constructively points out that effective use of GenAI in schools requires slow, steady work for any real progress. Its recommendations focused on reenvisioning what students learn, supporting educators in redesigning assignments, and investing in research, are all worthwhile considerations. Unfortunately, their generic nature limits their usefulness.

At a minimum, and going forward, useful guidance must clearly articulate what is meant by learning and how AI-powered tools might contribute to and detract from such learning. Proposals for AI use must account for the social, collaborative nature of learning; in other words, it must capture what occurs in a real classroom. Significantly, recommendations around AI *must* address privacy concerns and attempt to mitigate the potential and known harms associated with the use of GenAI, particularly when used with children and youth. Finally, recommendations need to account for a historical pattern of racial bias and an intentional vision for the use of these tools in a democratic society. Unfortunately, this report does not engage these topics and while it restates well-intentioned aspirations around using AI, it fails to provide actionable guidance for policymakers.

Notes and References

1 Reich, J. & Dukes, J. (2024). *Toward a new theory of arrival technologies: The case of ChatGPT and the future of education technology after adoption*. Unpublished manuscript. Retrieved September 28, 2025, from https://osf.io/preprints/edarxiv/x6vn7_v1

2 Young, J.R., Bevan, D., & Sanders, M. (2024). How productive is the productive struggle? Lessons learned from a scoping review. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 12*(2), 470-495. Retrieved September 28, 2025, from https://doi.org/10.46328/ijemst.3364

3 Kulesa, A.C., Mission, M., Croft, M., & Wells, M.K. (2025, June). *Productive struggle: How artificial intelligence is changing learning, effort, and youth development in education* (p. 11). Bellwether. Retrieved September 26, 2025, from https://bellwether.org/publications/productive-struggle/?activeTab=1

4 The report does not explain how it narrowed down the literature to these four components. While the report indicates that "these components have been identified by psychologists," there is no reference to relevant review articles.

5 Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178. Retrieved September 28, 2025, from http://www.jstor.org/stable/1466837

6 Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *The Journal of the Learning Sciences*, *18*(3), 327–369. Retrieved September 28, 2025, from https://www.tandfonline.com/doi/full/10.1080/10508400903013470

7 Sengupta-Irving, T. & Agarwal, P. (2017). Conceptualizing perseverance in problem solving as collective enterprise. *Mathematical Thinking and Learning*, *19*(2), 115–138. Retrieved September 28, 2025, from https://doi.org/10.1080/10986065.2017.1295417

8 Langer-Osuna, J.M. (2016). The social construction of authority among peers and its implications for collaborative mathematics problem solving. *Mathematical Thinking and Learning*, *18*(2), 107–124. Retrieved September 28, 2025, from https://www.tandfonline.com/doi/full/10.1080/10986065.2016.1148529

9 For example, see:

Takeuchi, M.A. (2016). Friendships and group work in linguistically diverse mathematics classrooms: Opportunities to learn for English language learners. *Journal of the Learning Sciences*, *25*(3), 411–437. Retrieved September 28, 2025, from https://doi.org/10.1080/10508406.2016.1169422

Wortham, S. E. (2006). *Learning identity: The joint emergence of social identification and academic learning*. New York, NY: Cambridge University Press.

10 Philip, T.M. & Garcia, A. (2013). The importance of still teaching the iGeneration: New technologies and the centrality of pedagogy. *Harvard Educational Review, 83*(2), 300-319.

11 Philip, T.M. & Sengupta, P. (2020). Theories of learning as theories of society. A contrapuntal approach to expanding disciplinary authenticity in computing. *Journal of the Learning Sciences,*

30(2), 330-349.

12 Chang, M.A., Tissenbaum, M., Philip, T.M., & D'Mello, S.K. (2025). Co-designing AI with youth partners: Enabling ideal classroom relationships through a novel AI relational privacy ethical framework. *Computers and Education: Artificial Intelligence*, 100364.

13 Crooks, R.N. (2024). *Access Is capture: How edtech reproduces racial inequality* (1st ed.). University of California Press. Retrieved September 28, 2025, from https://doi.org/10.2307/jj.18703062

14 Irby, B., Brown, G.H., Lara-Aiecio, R., & Jackson, S.A. (Eds.). (2013). *The handbook of educational theories*. IAP.

15 Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI can harm learning*. The Wharton School Research Paper. Retrieved September 28, 2025, from https://ssrn.com/abstract=4895486

16 Kulesa, A.C., Mission, M., Croft, M., & Wells, M.K. (2025, June). *Productive struggle: How artificial intelligence is changing learning, effort, and youth development in education* (p. 14). Bellwether. Retrieved September 26, 2025, from https://bellwether.org/publications/productive-struggle/?activeTab=1

17 Horwitz, J. (2025, August 14). Meta's AI rules have let bots hold 'sensual' chats with kids, offer false medical info. *Reuters*. Retrieved September 26, 2025, from https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/

18 Edouard, K. (2024). Black prompts not Included: AI bias impeding real-time co-creative ideation in Black students' conceptual design. In R. Lindgren, R., T.I. Asino, E.A. Kyza, C.K. Looi, D.T. Keifert, & E. Suárez. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 1366-1369). International Society of the Learning Sciences.

19 Hofmann, V., Kalluri, P.R., Jurafsky, D., & King, S. (2024). Dialect prejudice predicts AI decisions about people's character, employability, and criminality. *arXiv preprint*. Retrieved September 28, 2025, from https://arxiv.org/abs/2403.00742