Nexus for STEM Problem Solving and Transfer Research: Instruction First or Productive Failure First?[1]

Michael J. Jacobson

The University of Sydney

https:// ORCID 0000-0001-6521-9757

John Vulic

University of Technology Sydney

James A. Levin

University of California, San Diego

February 28, 2024

Abstract: Helping students develop deep conceptual knowledge and ways of thinking for knowledge transfer—that is, applying and using knowledge and problem-solving skills in new contexts and situations—are important goals for STEM literacy and for educational subjects in general. However, there is an ongoing research debate about whether learning designs and instructional approaches that provide *instruction first* or *Productive Failure first* are best for achieving these goals. This chapter provides an overview of this debate and of recent research findings challenging traditional views about the necessity of initially providing instruction. Studies are discussed (many in STEM subject areas) in which students first engage in problem solving and failure followed by instruction (i.e., Productive Failure [PF]) that have findings of superior conceptual learning and transfer by a factor of three compared to instruction first. As an example of STEM instruction involving PF, a study of students learning about complex systems and climate change using computer models is discussed that found significantly higher problem solving and transfer outcomes for the PF instructional approach compared to instruction first. The chapter concludes with considerations of future STEM learning research and the potential implications of innovative learning designs such as PF for enhancing students' ways of thinking and problem solving in science and related areas.

[1] Please do not quote from this manuscript without written permission from the first author. Suggested reference: Jacobson, M. J., Vulic, J., & Levin, J. A. (in press). Nexus for STEM problem solving and transfer research: Instruction first or Productive Failure first? In L. D. English & T. Lehmann (Eds.), *Ways of Thinking in STEM-based Problem Solving: Teaching and Learning in a New Era*. Routledge.

It is generally recognized that there are important general characteristics of STEM literacy such as students developing solid conceptual knowledge and ways of thinking about science and mathematics as well as the ability to use or transfer STEM knowledge to solve problems in new contexts and situations. However, these characteristics, particularly the ability to transfer knowledge, are challenging to achieve. Reasons for this include the Goldstone and Day (2012) observation that "schools often measure the efficiency of learning in terms of speed and retention of knowledge" (p. 149), with less attention to students generalizing and applying their acquired knowledge. Further, they argue there is an urgent need for "specific, validated techniques for teaching with an eye toward facilitating students' transfer of their learning" (p. 149). We concur with this view, but the question must be asked: How best to address this need?

In this chapter, we argue that a principled answer to this question requires engaging a current core debate or "fault line" (diSessa, 2006) in the fields of educational research and the learning and cognitive sciences. Prior "fault lines" include the cognitive/situative debate (Anderson et al., 2000; Jacobson et al., 2016), the knowledge-in-pieces/coherent-knowledge debate (diSessa, 2006; Slotta, 2011), and the quantitative/qualitative methods debate (Firestone, 1987). Each of these debates involved strong disagreements about theorizing, empirical findings, or methodological approaches—often in combination. These debates also persisted in the literature for years and even decades during which the research communities could not vindicate one camp or the other before a reconciliation was generally articulated and accepted (for a detailed discussion of these dynamics in the cognitive-situative debate, see Jacobson et al. (2016)).

Perhaps the most critical current educational research debate began after the start of the 21st century from the argument that superior learning occurs by providing instructional guidance (i.e., direct instruction [DI]) compared to "minimally guided" instructional approaches such as problem-based learning [PBL], inquiry, and discovery learning (Kirschner et al., 2006). This debate has evolved in the literature over the past nearly 20 years and we propose it may now be framed as the *instruction first* versus *Productive Failure (PF) first* debate. A large amount of empirical research—much of it involving STEM subjects—has been done into various facets of this debate, with the focus on students' learning outcomes based on these different types of learning designs and instructional approaches. We believe the current cumulative body of empirical findings challenges the Kirschner and associates' argument.

We first provide an overview of the *instruction first* or *PF first* debate, followed by a more detailed discussion of one of the STEM studies included in the Sinha and Kapur (2021) meta-analysis to illustrate a PF computer model-based learning design for students to learn about the scientific complexity of climate change and to discuss the significant learning and transfer findings compared to instruction first. The chapter concludes with considerations of directions for future research and the potential implications of STEM teaching and policy.

The Instructional Guidance Debate

The Superiority of Direct Instruction?

This debate was initially framed from arguments made in Kirschner et al. (2006) based on disputes about how instructional guidance can impact learning since the mid-twentieth century. They proposed that students best learn specific concepts and procedures in different subjects from direct instructional guidance and *not* from learning activities in minimally or unguided environments. Direct instructional guidance (i.e., direct instruction; we will use these two terms interchangeably in this chapter) is defined as the use of cognitively appropriate learning strategies to fully explain concepts and procedures to students, with learning viewed as changes in long-term memory. Rather than defining minimally guided instruction, they list several instructional approaches they regard as exemplars, including discovery learning, problem-based learning (PBL), inquiry learning, experiential learning, and constructivist learning. They claim these are "differently named but essentially pedagogically equivalent approaches (p. 75)" in which students are involved with learning activities in which they personally discover or construct understandings about concepts and procedures. Kirschner and colleagues discuss several studies that compared guided and unguided/minimally guided instruction. They assert the research evidence conclusively supports the superiority of guided instruction for initial to intermediate level student learning, which they believe is consistent with research about expert-novice difference and with cognitive load theory.

In a response to Kirschner et al. (2006), Hmelo-Silver et al. (2007) disputed the Kirschner et al. claim about "pedagogically equivalent approaches" and asserted they had conflated learning approaches such as problem-based learning (PBL) and inquiry learning (IL) with unguided instruction. Unlike approaches such as discovery learning, Hmelo-Silver et al. argue that PBL and IL make extensive use of scaffolding, which is a type of instructional guidance. This use of scaffolding in turn reduces cognitive load and facilitates learning in complex domains, as well as addressing other educational goals such as learning epistemic practices and skills such as self-directed learning and collaboration. However, while the argument is persuasive that Kirshner et al. incorrectly categorized approaches such as PBL and IL and being "minimally guided," Hmelo-Silver et al. do not directly challenge the core assertion of Kirschner and colleagues, which is direct instructional guidance results in superior learning outcomes compared to other instructional approaches.

Researching, Reframing, and (Possibly) Resolving the Debate

During the decade and a half following the publications of Kirschner et al. (2006) and Hmelo-Silver et al. (2007), a body of empirical learning research has been accumulating that directly or indirectly investigates various aspects of the assertion made by Kirschner et al. (2006) that direct instructional guidance approaches are superior for learning compared to minimally guided instruction approaches. These studies have investigated a range of issues that provide a more granular understanding of design considerations for learning and instruction, such as specific details of various learning designs, wider range of subjects (although the majority of studies were focused predominately within STEM fields), students' grade level, assessments for a greater range of knowledge variables (e.g., procedural, conceptual, transfer), and considerations of the social contexts of learning.

Of special relevance to research related to the debate during this period was the articulation of a new type of learning design—*Productive Failure* (PF) (Kapur & Bielaczyc, 2012) that differed in key ways from the designs reviewed by Kirschner et al. (2006). Unlike the single phase of a learning intervention generally found in the research on direct instruction/minimally guided instruction approaches reviewed by Kirschner et al., PF employs two phases. First, in the Exploration and Generation (or just Exploration) phase, students work on a problem activity for which they, as novices, are likely to have incomplete or incorrect knowledge and thus would likely struggle (i.e., fail) in reaching a canonical or "correct" solution. Second, in the Consolidation and Knowledge Assembly (or Consolidation) phase, the teacher provides instruction that builds on the students' ideas and provides the concepts and procedures necessary to solve the problem. Kapur and Bielaczyc (2012) theorize that students *activate prior knowledge* in the Exploration phase, and then in the Consolidation phase they *contrast and compare* their inaccurate or partially correct solutions to the teachers' concepts and procedures, resulting in the assembly of new knowledge. The early research on PF generally used direct instruction approaches as a control condition, and found significantly higher gains in terms of conceptual knowledge and transfer for PF when compared to DI (Kapur, 2006, 2010, 2014; Kapur & Bielaczyc, 2012).

The language used to describe the debate also has evolved. Jacobson et al. (2015) proposed that the Kirschner et al. (2006) framework of "direct instructional guidance" and "minimally guided instruction" could be reframed as "high pedagogical structure" and "low pedagogical structure." They analyzed the empirical studies discussed by Kirschner et al. as comparisons between high pedagogical structure (i.e., direct instruction) and low pedagogical structure (i.e., minimally guided instruction). Further, they noted that no discussion was provided by Kirshner et al. of studies in which there were *different sequences of pedagogical structure* in the learning activities, such as Schwartz and Bransford (1998), VanLehn et al. (2003), Bjork and Linn (2006), and Kapur (2006).

Loibl et al. (2017) proposed a different terminology to describe learning designs with two phases: "problem-solving followed by instruction" (PS-I) and "instruction followed by problem-solving" (I-PS). They also reviewed 34 studies of PS-I approaches selected from 20 published articles, with a focus on PF (Kapur & Bielaczyc, 2012) research and Invent with Contrasting Cases (Schwartz & Martin, 2004) as subsets of PS-I. Their review found significant learning outcomes from PS-I approaches compared to I-PS when specific design features were used, such as instruction that built on students' ideas and providing contrasting cases during problem solving. They also proposed a set of cognitive mechanisms associated with positive PS-I learning outcomes, such as activation of student's prior knowledge, awareness of knowledge gaps, and recognition of deep features (not just surface features) of a problem.

Further developments in the instructional guidance debate, such as research reviews by Darabi et al. (2018) and Sinha and Kapur (2019), expanded the range of targeted concepts for learning and the geographical locations where the research was conducted. In the most detailed and extensive meta-analysis of this body of research to date, Sinha and Kapur (2021) included 53 studies with 166 comparisons of PS-I and I-PS designs. Their overall analysis found PS-I had a significant and large Hedge's g effect size of 0.36 over I-PS. Further, for PS-I instructional approaches that implemented the design principles of Productive Failure (PF), there was a Hedge's g of 0.58 over I-PS. To contextualize these findings, Kraft (2020) has proposed baseline benchmarks for effect sizes in educational interventions, categorizing them as small (less than 0.05), medium (0.05 to 0.20), and large (0.20 or greater). Thus, the PS-I versus I-PS effect size of 0.36 is nearly double and the PF versus I-PS effect size of 0.58, which is nearly triple what is considered a large effect for educational interventions.

The Sinha and Kapur (2021) analysis also provided a more granular insight into the learning efficacy associated with different learning design elements than in earlier reviews. As a metric to distinguish PF from PS-I, they proposed a *PF fidelity score* that includes criteria such as eliciting multiple representations and subject matter ideas from students, engagement in group work, instruction building on students' solutions, and a social facilitation element in the instruction phase. They found that the studies with high PF fidelity scores yielded the highest effect size results in their posttest assessments, indicating a strong positive correlation between PF fidelity and educational outcomes.

The main results of the Sinha and Kapur (2021) meta-analysis found that PF was best in terms of learning gains—and both PF and PS-I were superior—when compared to I-PS for more cognitively challenging conceptual knowledge. It was also found that PS-I, PF, and I-PS approaches had statistically equivalent gains in learning procedural knowledge. Data analysis revealed a nuanced pattern, where I-PS was found to be more effective with younger students in grades two to five and for learning domain general skills, which was probably due to a lack of prior knowledge across the younger cohort.

To summarize, nearly two decades after Kirschner et al. (2006) made the provocative claim of the superiority of direct instructional guidance approaches, the preponderance of empirical data does *not* support their assertion about the superiority of instruction first for learning. Rather, the research to date demonstrates that *productive failure first* has significant advantages over direct instruction or I-PS approaches for fostering deeper conceptual understandings as well as for developing transferrable problem-solving skills and equivalent learning as I-PS for procedural knowledge.

Productive Failure for Model-based Learning about Complex Systems and Climate Change

In this section, we illustrate a PF intervention involving STEM learning and problem solving from the Jacobson et al. (2017) study, which was included in the Sinha and Kapur (2021) meta-analysis. Before doing so, we observe that nearly all of the PS-I/I-PS studies reviewed by Sinha and Kapur (2021) involved learning activities in conventional classroom settings where students verbally discussed problems—sometimes with hand written notes, diagrams, or mathematical calculations—and teachers verbally provided instruction.

In contrast, the Jacobson et al. (2017) study was the only intervention included in Sinha and Kapur (2021) to use problem solving activities with computational models. Part of the context for Jacobson et al. (2017) was a recognition that scientists increasingly employ computer modeling and scientific visualization tools to study a wide range of physical and social phenomena (Epstein, 2006; Jacobson & Wilensky, 2022; Mitchell, 2009) and also recognition for research demonstrating significant learning outcomes from model-based approaches in STEM areas such as physics, chemistry, genetics and evolution, and engineering (Blikstein & Wilensky, 2010; Goldstone & Wilensky, 2008; Horwitz et al., 2010; Jacobson et al., 2015; Sengupta & Wilensky, 2009). However, few of the earlier model-based learning studies provided specific details about instructional sequencing, and most seem likely to be categorized as I-PS where the model-based activities were provided after initial instruction about the targeted science topics.

The Jacobson et al. (2017) study involved Australian secondary students in four classes who used NetLogo (Wilensky, 1999) computer models to learn core concepts about complex systems (typically not formally taught until master's level science courses) and climate change ideas found in secondary level science courses. Table 1 shows the daily topics and content knowledge and Figures 1 and 2 provide screen shots of a complex systems forest fire model and a climate change model.

<Insert Table 1 approximately here.>

<Insert Figures 1 and 2 approximately here.>

The study involved a pretest, the experimental and comparison learning interventions over four class periods, and a posttest; see Table 2 for the daily components of the study. In the PF experimental condition that involved two classes, students first worked collaboratively on problems with the computer models (i.e., the PF Exploration and Generation phase). The teacher then provided instruction (i.e., the PF Consolidation and Knowledge Assembly phase) that explained the complex systems and climate concepts relevant for solving the problems, after which the student worked on a second set of problems for which solutions required the use of the complex systems and climate concepts explained in the instruction. The Direct Instruction (DI) comparison group (i.e., I-PS) also involved two different classes and used the same computer models and problem activities, but with the teacher instruction provided before the students worked on the two sets of problems.

<Insert Table 2 approximately here.>

The dependent measures for the learning outcomes were: (a) declarative knowledge, (b) conceptual knowledge (Coleman, 1998), (c) near transfer, and (d) far transfer. Items for assessing declarative and conceptual knowledge had two parts, such as: "(a) What are examples of emergent properties in climate systems? (b) Please explain." Students were intended to provide factually oriented information (i.e., "knowing what"; Bransford et al., 2000) in part (a) and conceptual knowledge (i.e., "knowing how or why" or explanatory information Coleman, 1998)) in part (b). Near transfer was assessed with the "Butterfly Effect" Problem: *It has been said that a butterfly flapping its wings in Brazil can jiggle the air and thus can help cause a snowstorm in Alaska. Is this possible? If so, how? If not, why not?* This is regarded as "near transfer" problem because a reasonably complete expert explanation would provide relevant declarative and conceptual domain knowledge about climate and complex systems from the instruction, but this specific problem is new (i.e., not worked on in the intervention) to the students. Far transfer (sometimes called far across domain transfer) was assessed with the "Robot Mining" problem (Resnick, 1994): *How can autonomous robots on a remote planet effectively and efficiently mine for gold?* This is a far transfer problem as it is about a completely different domain than those used in the experimental learning interventions (i.e., robotic mining versus biological/environmental sciences) and requires using specific complex systems concepts (e.g., agent interactions, positive feedback, self-organization, emergent properties) to solve the problem.

The qualitative observations of the classes noted that the PF classroom environment was lively as the students collaboratively worked on the daily problems and discussed their ideas while using the computer models. This seemed to indicate that the students were engaged and motivated as they were generating ideas while using the models, which was the intention of the

PF Generation and Exploration phase. During the PF Consolidation and Knowledge Assembly phase, the teacher used an interactive white board to run the computer models and to explain the targeted climate and complex systems ideas for each day. The students often would ask questions and try out computer model parameter settings the teacher was using on their own laptop computers. In contrast, the DI/I-PS classroom environment was very quiet. As in a typical teacher led class, the teacher presented the climate and complex systems content and the students silently took notes, rarely asking questions. During the two problem solving activities, the students mainly looked at the notes taken during the teacher's instruction rather than collaboratively talking or reasoning with the computer models to answer the problems. These qualitative findings are consistent with a key design feature of PF compared to DI/I-PS, which is the importance of the "social surround" (Kapur & Bielaczyc, 2012).

For the quantitative results, there were no significant differences between the PF and DI groups on the pretest. On the posttest, no significant differences were found on the declarative posttest scores between the PF and DI treatment conditions, however, the PF group had significantly higher results on the posttest conceptual knowledge and transfer scores, as shown in Figure 3. The Hedge's *g* effect size for these results was 0.95 (Kapur, 2019, Aug). This is nearly five times greater than the 0.20 Kraft (2020) benchmark of a large effect size, and also larger than the PS-I versus I-PS effect size of 0.36 and the PF versus I-PS effect size of 0.58 reported in Sinha and Kapur (2021).

There are three main implications of the Jacobson et al. (2017) findings. First, a limitation in much current research involving model-based learning is the assumption that merely using the technology by students will automatically enhance their learning. This assumption is clearly challenged by these results in which an I-PS design significantly attenuated the efficacy of learning with computer models compared to a PF design. Future research is needed to further explore the issue of which learning designs are optimal for fostering STEM learning from computational modeling and related systems. Second, given the increasing use of computational modeling and simulations in 21st century science, future research could also investigate potential advantages for students to learn scientific knowledge and problem-solving skills *with* computer modeling and related systems, rather than learning the knowledge and skills separately and then later "applying" (or transferring) the new knowledge in computational models. Third, and perhaps most important for the "real world" of classroom teaching, is the finding that students using computer models in a PF approach seemed genuinely engaged and motivated compared to students in a DI setting using the same computer models and problem-solving activities. Given many students find STEM classes uninteresting, further research is clearly warranted into the potential motivational aspects of using PF with computer modeling for learning science and related knowledge.

<Insert Figure 3 approximately here.>

Conclusion

This chapter has discussed research involving different instructional approaches for helping students learn conceptual knowledge and problem-solving skills for STEM literacy and as well as general educational subjects. An overview is provided of the contention by Kirschner et al. (2006) that initially provided direct instructional guidance or I-PS approaches are superior to other instructional approaches, which we refer to as the *instruction first* or *Productive Failure* *first* debate. The current status of the research related to this debate is probably best summarized in the Sinha and Kapur (2021) meta-analysis, which found PS-I learning designs, in particular those in which students first engaged in problem solving and failure followed by instruction (i.e., Productive Failure [PF]), had significantly higher effect sizes for conceptual learning and transfer compared to I-PS approaches. As an example of STEM instruction involving PF, a study by Jacobson et al. (2017) is discussed that used PF for learning about complex systems and climate change with computer models, which found superior learning outcomes when compared with direct instruction provided first before using the same computer models.

Overall, the cumulative evidence from a wide range of empirical studies shows that students can more deeply learn STEM knowledge and more successfully use or transfer this knowledge in problem solving. Future research in this area is warranted for longer duration studies to validate (or not) PF and other PS-I approaches as well as to better understand their possible influence on longitudinal trajectories of learning, such as work on preparation for future learning (Bransford & Schwartz, 1999; Sinha & Kapur, 2021). In closing, it is hoped this chapter will help stimulate interest in the potential implications of innovative learning designs such as PF and the use of computational modeling for problem solving activities to greatly enhance students' ways of thinking and problem solving in—and enjoyment of—science and related subjects.

Acknowledgement

The research discussed in this chapter was funded in part by grants to the first author from the Australian Research Council (ARC) programs DP150102144, DP150104267, and LP100100594, and the New South Wales Department of Education and Communities. Also, the feedback from a reviewer on an earlier version of this manuscript is gratefully acknowledged.

References

Anderson, J. R., Greeno, J. G., Reder, L. M., & Simon, H. a. (2000). Perspectives on Learning, Thinking, and Activity. *Educational Researcher*, *29*(4), 11-13. https://doi.org/10.3102/0013189X029004011

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science: Introducing desirable difficulties. *APS Observer*, *19*(29), 39-39.

Blikstein, P., & Wilensky, U. (2010). MaterialSim: A Constructionist Agent-Based Modeling Approach to Engineering Education. In *Designs for Learning Environments of the Future* (pp. 17-60). https://doi.org/10.1007/978-0-387-88279-6_2

Bransford, J. D., Brown, A. L., Cocking, R. R., & Donovan, S. (2000). How people learn: Brain, mind, experience, and school (expanded edition). In (pp. 357-357). Washington, DC: National Academy Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Hejad & P. D. Pearson (Eds.), (pp. 61-100). American Educational Research Association.

Coleman, E. (1998). Using Explanatory Knowledge During Collaborative Problem Solving in Science. *Journal of the Learning Sciences*, *7*(3), 387-427. https://doi.org/10.1207/s15327809jls0703&4_5

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: a meta-analysis of the empirical studies. *Educational Technology Research and Development*, *66*(5), 1101-1118. https://doi.org/10.1007/s11423-018-9579-9

diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In R. K. Sawyer (Ed.), (pp. 265-281). Cambridge University Press.

Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.

Firestone, W. A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, *16*(7), 16-21.

Goldstone, R. L., & Day, S. B. (2012). Introduction to New Conceptualizations of Transfer of Learning. *Educational Psychologist*, *47*(3), 149-152. https://doi.org/10.1080/00461520.2012.695710

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer through complex systems principles. *Journal of the Learning Sciences*, *17*(4), 465-516.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107. https://doi.org/10.1080/00461520701263368

Horwitz, P., Gobert, J. D., Buckley, B. C., & O'Dwyer, L. M. (2010). Learning genetics with dragons: From computer-Based manipulatives to hypermodels. In M. J. Jacobson & P. Reimann (Eds.), (pp. 61-87). Springer-Verlag.

Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Towards a complex systems conceptual framework of learning. *Educational Psychologist*, *51*(2), 210-218. https://doi.org/10.1080/00461520.2016.1166963

Jacobson, M. J., Kim, B., Pathak, S., & Zhang, B. (2015). To guide or not to guide: issues in the sequencing of pedagogical structure in computational model-based learning. *Interactive Learning Environments*, *23*(6), 715-730. https://doi.org/10.1080/10494820.2013.792845

Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction*, *52*, 1-14. https://doi.org/10.1016/j.learninstruc.2017.03.007

Jacobson, M. J., & Wilensky, U. (2022). Complex Systems and the Learning Sciences: Educational, Theoretical, and Methodological Implications. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed.). Cambridge University Press.

Kapur, M. (2006). Productive failure. *Cognition and Instruction*, *26*(3), 307-313.

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, *38*(6), 523-550. https://doi.org/10.1007/s11251-009-9093-x

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008-1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2019, Aug). *When productive failure fails: Keynote at the* Annual meeting of the European Association for Research on Learning and Instruction (EARLI), Aachen, Germany.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences*, *21*(1), 45-83. https://doi.org/10.1080/10508406.2011.591717

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241-253. https://doi.org/10.3102/0013189x20912798

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review*, *29*(4), 693-715. https://doi.org/10.1007/s10648-016-9379-x

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Resnick, M. (1994). *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*. MIT Press.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475-522.

Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction*, *22*(2), 129-184. https://doi.org/10.1207/s1532690xci2202_1

Sengupta, P., & Wilensky, U. (2009). Learning electricity with NIELS: Thinking with electrons and thinking in levels. *International Journal of Computers for Mathematical Learning*, *14*(1), 21-50, DOI: 10.1007/s10758-10009-19144-z.

Sinha, T., & Kapur, M. (2019). When Productive Failure Fails. 41st Annual Conference of the Cognitive Science Society, Montreal, Canada.

Sinha, T., & Kapur, M. (2021). When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. *Review of Educational Research*, *91*(5), 761-798. https://doi.org/10.3102/00346543211019105

Slotta, J. D. (2011). In Defense of Chi's Ontological Incompatibility Hypothesis. *Journal of the Learning Sciences*, *20*(1), 151-162. https://doi.org/10.1080/10508406.2011.535691

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *2*(3), 209-249.

Wilensky, U. (1999). *NetLogo*. Center for Connected Learning and Computer-Based Modeling. Northwestern University (http://ccl.northwestern.edu/netlogo).

Tables

Table 1

Daily topics and content knowledge

Day	Complex Systems Topics	Concepts for Climate Model	Concepts for Complex Systems Models
0	Pre-test		
1	Feedback in dynamic systems	Climate model with feedback • Global temperature, greenhouse effect, cloud cover	Ants foraging model • Input, output, positive feedback
2	Emergence	Wind and storm model • Convection, wind, greenhouse effect, enhance greenhouse effect	Birds flocking model • Emergence, micro level of systems, macro level of systems
3	Dynamic equilibrium	Carbon cycle model • Equivalent carbon, caron cycle, fossil fuel, carbon sinks	Wolf-sheep predation model • Equilibrium, dynamic equilibrium, close system
4	Tipping points and positive feedback	Climate model with water feedback • Atmospheric water feedback, albedo	Forest fire model • Positive feedback, tipping points, linear versus non-linear effects
5	Post-test		

Note: This table is from Jacobson et al. (2017, p. 5).

Table 2

Daily components of study

Direct Instruction (DI) Comparison Condition				
Direct Instruction Learning Design	Teacher Instruction	Practice		
Classroom Activities	Teacher Lecture	• Problem 1 • Complex Systems Model	• Problem 2 • Climate Model	• Problem 3 Climate and Complex Systems Models
Duration	Phase 1: 15-20 Minutes	Phase 2: 15-20 Minutes		Phase 3: 10-15 Minutes
Productive Failure (PF) Experimental Condition				
Productive Failure	Generation and Exploration		Consolidation and Knowledge Assembly	
Classroom Activities	• Problem 1 • Complex Systems Model	• Problem 2 • Climate Model	Teacher Lecture	• Problem 3 • Climate and Complex Systems Models
Duration	Phase 1: 15-20 Minutes		Phase 2: 15-20 Minutes	Phase 3: 10-15 Minutes

Note: This table is from Jacobson et al. (2017, p. 5).

Figures

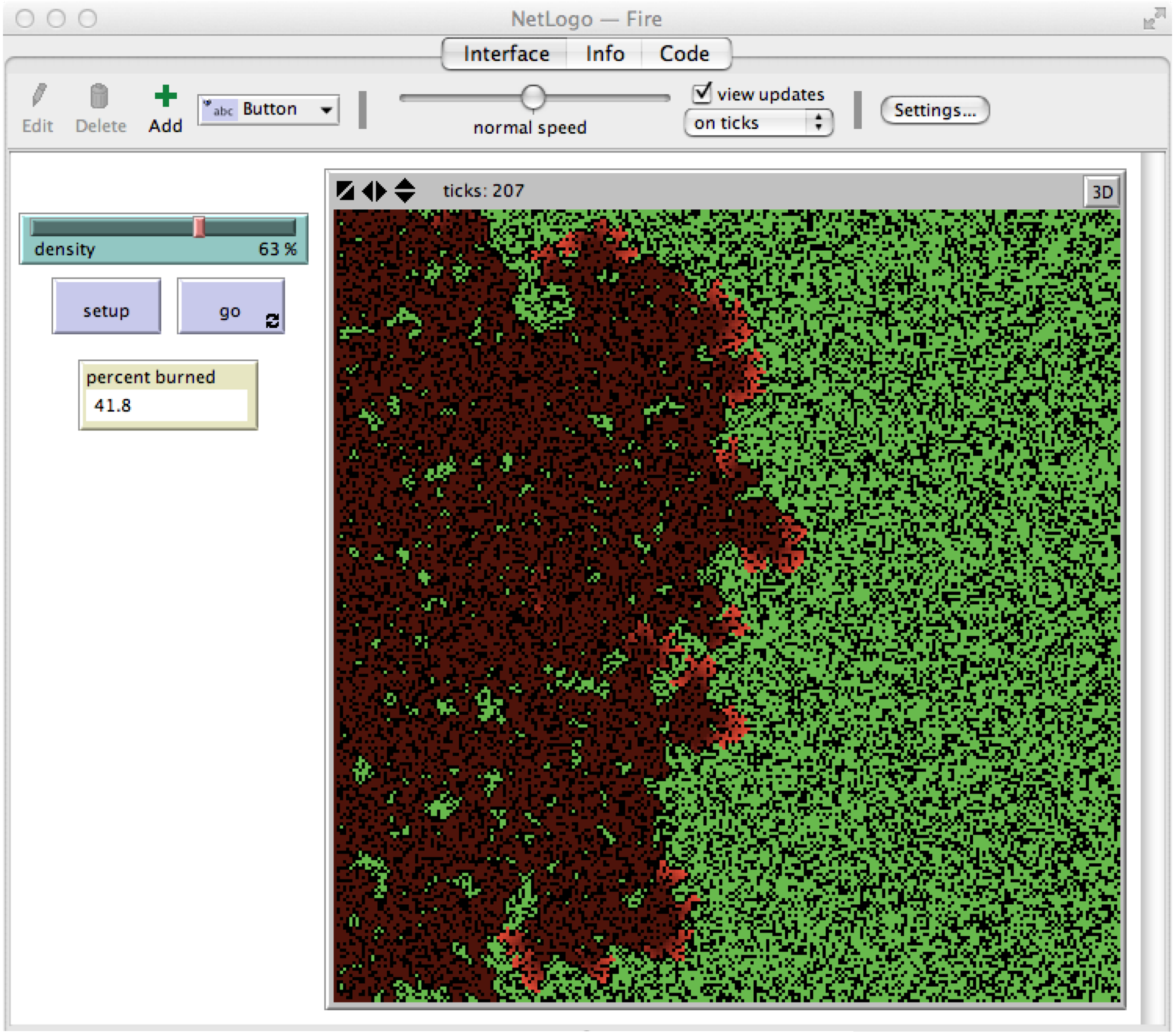

Figure 1. Screen shot of the NetLogo Forest Fire model.

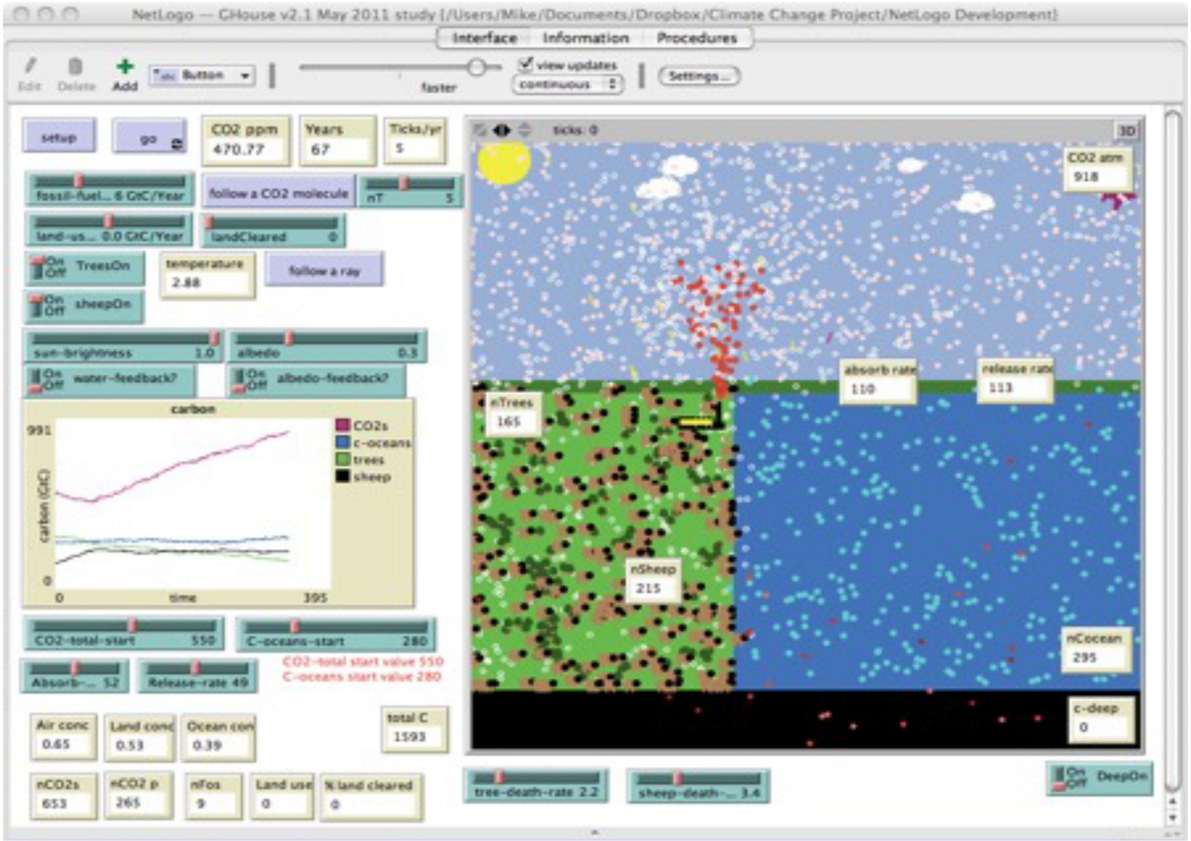

Figure 2. Screen shot of the Climate Model with Water Feedback.

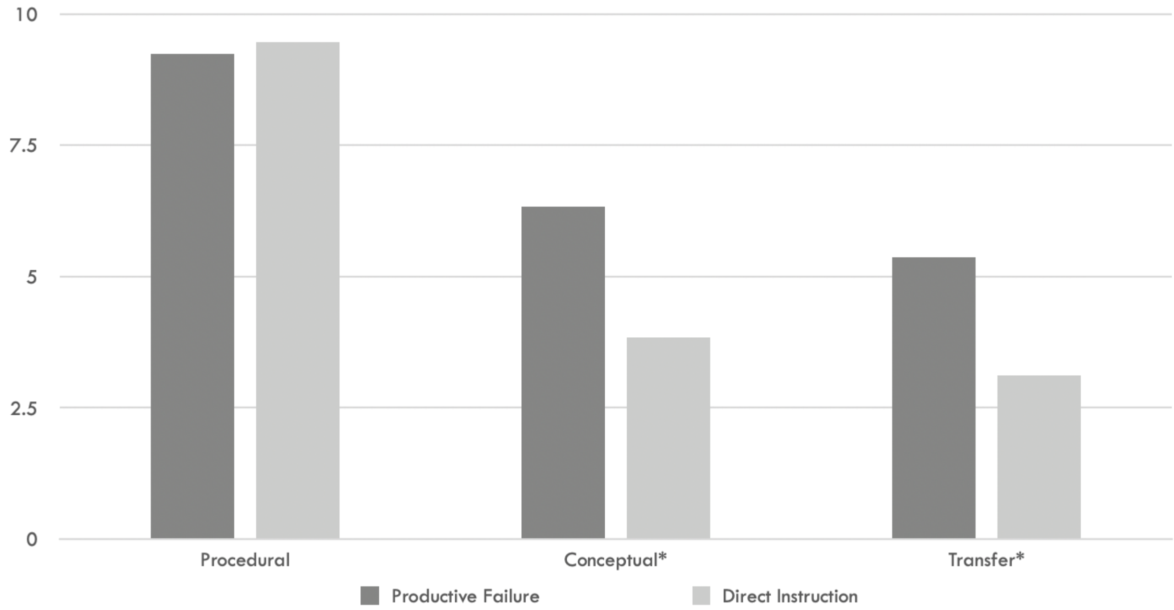

Figure 3. Normalized scores on the three types of knowledge assessed in Jacobson et al. (2017). An "" indicates a statistically significant difference.*