COMPUTATIONAL EDUCATIONAL RESEARCH: AN EXPLORATION OF INSTRUCTIONAL FRAMEWORKS

John Vulic, Michael J Jacobson, James A Levin
University of Technology Sydney, The University of Sydney, University of California, San Diego

Abstract: Our paper presents a research approach that conceptualizes education as a complex adaptive system, investigated using methods from computational science. We describe the development of agent-based models that simulate learning processes within classroom environments, drawing on existing quantitative and qualitative data. These models support computational experiments that are iteratively refined using real-world classroom data, enabling exploration of questions such as: "Which pedagogical approaches most effectively support student learning and transfer?" Findings from our simulations, based on data comparing instructional approaches, suggest that Productive Failure may be particularly effective in fostering deeper learning and transfer. We conclude by discussing implications for the design and use of computational modeling in educational research.

Conceptual Rationale and Aims

Our paper presents a computer modeling approach that responds to recent calls for advancing computational educational research (Williamson, Potter, & Eynon, 2019). We use computational modeling to represent facets of complex educational systems (Jacobson, 2020), in which learning dynamics unfold across multiple system levels—such as individual cognition, collaborative learning, and the broader social contexts of schools—and exhibit emergent phenomena, including individual learning outcomes and educational equity. Our primary objective is to illustrate the development of computer models of classroom learning grounded in existing data and to conduct computational experiments to evaluate whether the models adequately reflect observed outcomes. Where discrepancies arise, we identify areas for further investigation and refinement. We conclude with a discussion of how computational modeling can inform the design of real-world classroom experiments and contribute to educational research more broadly.

Literature Review

Jacobson, Levin, and Kapur (2019) have proposed that computational science techniques—long applied to the study of physical, biological, and social complex systems—can also complement traditional quantitative and qualitative methods in educational research. Computational approaches offer the capacity to model complex, dynamic systems that evolve over time, capturing emergent phenomena and interactions across multiple levels. While quantitative methods are valuable for testing hypotheses and identifying statistically significant relationships between variables, they may oversimplify complex phenomena by focusing on predetermined variables and generalizations, often overlooking contextual nuance, local perspectives, and the subjective decisions embedded in the research process (Johnson & Onwuegbuzie, 2004). In contrast, qualitative methods offer rich, contextual insights but may be limited by subjectivity, reduced generalizability, time-intensive data collection, challenges in replication, complex analysis procedures, and ethical considerations (Denzin & Lincoln, 2018, pp. 1–15). Our research seeks to move beyond the quantitative–qualitative paradigm divide by advancing computational educational research as a methodological bridge—one that combines the strengths of both approaches to model and generate insights into complex learning environments (Jacobson, Levin, & Kapur, 2019).

Our research employs a multi-mediator modeling (MMM) approach (Levin & Datnow, 2012), which integrates elements of agent-based modeling (a "bottom-up," algorithm-based method; Wilensky & Rand, 2015) and system dynamics (a "top-down," equation-based method; Smith, 2007). We propose that computational approaches such as agent-based modeling, system dynamics, and multi-mediator modeling enable the simulation of interactions among multiple agents or factors over time, capturing emergent and non-linear behaviors that are difficult to observe directly (Jacobson, Kapur, & Reimann, 2016). These methods complement traditional quantitative and qualitative approaches by providing dynamic visualizations, allowing for the testing of hypothetical interventions, and offering the potential to inform real-time educational decision-making.

Our study employs computational educational research to compare pedagogical approaches using real-world data, with the goal of generating scalable, evidence-based insights that more effectively capture the complexity of classroom learning than traditional methods. Our initial analysis focuses on two contrasting instructional designs. The first is Productive Failure (PF) (Kapur, 2008), an approach in which learners engage with complex, ill-structured problems without immediate support, often experiencing initial failure. This productive struggle is intentionally designed to activate prior knowledge and cognitive engagement, ultimately leading to deeper conceptual understanding and improved transfer of learning when followed by explicit instruction. The second approach is Direct Instruction (DI), as described in the structured teaching model proposed by Engelmann and Carnine (1982), which is grounded in the belief that all students can learn through carefully sequenced, explicit instruction that emphasizes clarity, mastery, and the systematic development of generalizable skills. Of particular interest in our study is the sequencing of DI followed by student-centered problem-solving.

In their meta-analytic review of 53 empirical studies on different pedagogical approaches, Sinha and Kapur (2021) focused on three instructional methods: Instruction followed by Problem Solving (I-PS), commonly referred to as "regular teaching" or direct instruction; Problem Solving followed by Instruction (PS-I), often associated with problem-based or collaborative learning; and Productive Failure (PF). Their findings revealed that PS-I produced superior outcomes compared to I-PS, with a Hedges' *g* effect size of 0.36—representing a 1.8 times improvement over regular teaching (Hedges, 1981). Similarly, PF outperformed I-PS with a Hedges' *g* of 0.58, indicating a 2.9 times enhancement. These results were further supported by Jacobson et al. (2017), who reported a Hedges' *g* of 0.95 for PF compared to I-PS—an impressive 4.8 times advantage over traditional instruction. These findings underscore the potential of innovative pedagogical methods like PF to significantly enhance learning outcomes beyond conventional techniques. A central question addressed in this paper is whether such empirical findings can be replicated and better understood through computational educational research.

Method

The Learning Models developed for this project were built using the agent-based modeling tool NetLogo (Wilensky, 1999) (Figures 1-8). These models represent the educational problem space as a complex system (Jacobson et al., 2019), composed of nodes (i.e., circles) representing agents or elements, and links (i.e., lines) denoting their interactions. Two distinct models were created: one simulating the Productive Failure (PF) approach (view model) and the other simulating Direct Instruction (DI) (view model). Both models are run independently of each other.

The first level of each model represents the instructional phase, containing nodes that reflect the pedagogical fidelity of the instructional approach—either PF or DI—including the exploration and consolidation phases within a typical 60-minute classroom lesson for a PF lesson, and the instruction and problem-solving phases within a typical DI lesson. Subsequent levels are common to both models.

The second level captures student affect, specifically *student affective boost*, which reflects the range of students' emotional engagement, from low to high. Kapur (2024) defines affective boost as the motivational uplift students experience as they move toward a learning goal, influenced by factors such as the desire to avoid loss, recognition of progress, increasing motivation when near goal completion, and the satisfaction of task completion.

The third level represents the *student cognitive level*, modeling the potential for cognitive development during the lesson. This includes elements such as *prior knowledge*, *inert* and *elaborated schemas*.

External to these levels are nodes representing the dependent variables: *knowledge* and *transfer*, which serve as indicators of student learning and the transfer of learning.

In the visualizations, green lines indicate positive or supportive interactions (ranging from 0 to +1), while red lines denote negative or inhibitory interactions (ranging from -1 to 0). The models display numeric node values that reflect the cumulative effects of these network interactions across simulation runs.

Settings for the independent variables used in the representative computational experiments—adjusted via box sliders for each node—are displayed in each box slider in Figures 1-8. With each model run, the sizes of the dependent variable nodes and their corresponding numerical values change dynamically. Figures 1-8 illustrate the initial settings for the independent variables.

These variables include *Productive Failure (PF)* fidelity, which refers to the extent to which a learning activity adheres to the core design principles of the productive failure approach (Kapur, 2008). *Affective Boost (AB) Low* represents a composite of emotions, moods, interests, and motivational states that influence how students perceive and respond to learning situations (Kapur, 2024). For instance, a high AB Low setting indicates a significant reduction in student engagement or enthusiasm for a task.

Compatible prior knowledge refers to well-organized knowledge structures that can be readily accessed and applied to new situations, while *incompatible prior knowledge* consists of disconnected or poorly structured knowledge, making transfer and application more difficult (Bransford, 2000, p. 237). *Inert schema* relates to knowledge that learners possess but fail to apply in other contexts (Renkl, Mandl, & Gruber, 1996). Finally, *No Assembly* refers to fragmented knowledge structures in which students focus on superficial aspects of problems rather than underlying principles, thus limiting flexible knowledge application (Chi, Feltovich, & Glaser, 1981).

In this study, we conducted four computational experiments beginning from the initial Pretest state, altering only the pedagogical approach fidelity independent variable across both the PF and DI models. These experiments drew on data from the meta-analytic review by Sinha and Kapur (2021), as discussed above. Based on their findings, we assigned activation values to the PF Fidelity and DI Fidelity nodes in each model. A “small” node size corresponds to an activation value of 0, indicating the absence of instruction. An activation value of 0.20 represents the Hedges' *g* effect size for the I-PS condition (equivalent to approximately one year of instruction). A value of 0.36 reflects the PS-I Hedges' *g* effect size (approximately 2.8 years of instruction), while a value of 0.58 corresponds to the PF Hedges' *g* effect size (approximately 3.9 years of instruction).

Results

Table 1 presents the settings for the independent variables and the corresponding results of the computational model experiments for the dependent variables *Knowledge* and *Transfer*.

<INSERT TABLE 1 HERE>

Figures 1-8 show screenshots of the Learning Model link/node states for the pretest and each experiment for both PF and DI models. Instructional fidelity effect sizes for each model from Sinha & Kapur (2021).

<INSERT FIGURES 1-8 HERE>

Discussion

Before any instructional interventions were applied, a pretest was conducted on both models to establish baseline levels of knowledge and transfer for the PF and DI conditions (as shown in Figures 1 and 2). The instructional fidelity effect size was set to 0.00, indicating no instructional manipulation at this stage. As expected, when the models were run, both the PF and DI groups recorded zero gains in knowledge and transfer. Additionally, the *student affective boost* and *student cognitive level* elements remained unchanged across both models. This pretest confirmed that neither instructional approach yields any impact in the absence of instruction, establishing a consistent baseline across both models.

In the first experiment, the instructional fidelity effect size was modestly increased to 0.20 across both models (as shown in Figures 3 and 4), representing the introduction of a low level of structured instructional support. Under these conditions, the Productive Failure model (Figure 3) demonstrated a small improvement in knowledge acquisition (effect size = 0.08), while the Direct Instruction model (Figure 4) showed no measurable gains. Neither approach resulted in any improvement in transfer performance, suggesting that low instructional fidelity was insufficient to support deeper learning or the application of concepts beyond rote memorization.

With instructional fidelity further enhanced to an effect size of 0.36, Experiment 2 revealed more pronounced differences between the two instructional approaches. The Productive Failure model resulted in a moderate gain in knowledge (effect size = 0.49) and a small improvement in transfer (effect size = 0.16). In contrast, the Direct Instruction model showed only a minimal gain in knowledge (effect size = 0.05) and no improvement in transfer.

In Experiment 2 of the PF model (Figure 5), noticeable shifts emerged in the node configurations: there was an increase in *Affective Boost High* and *Elaborated Schema nodes*, along with corresponding decreases in *Affective Boost Low* and *Inert Schema* nodes. By comparison, in the Experiment 2 DI model (Figure 6), there was no increase in the *affective boost* and only a minimal decrease in the *No Assembly* node, with all other elements remaining unchanged. These results suggest that Productive Failure begins to demonstrate stronger benefits over Direct Instruction as the fidelity of instructional elements improves.

Experiment 3 represented the highest level of instructional fidelity, with an effect size of 0.58 (as shown in Figure 7). Under these conditions, the PF model produced substantial gains in both knowledge (effect size = 0.73) and transfer (effect size = 0.59), demonstrating its effectiveness when implemented with high fidelity. While the Direct Instruction model also showed an increase in knowledge (effect size = 0.25), it continued to yield no improvement in transfer performance (as shown in Figure 8).

In Experiment 3 of the PF model (Figure 7), we observe even more pronounced shifts in node sizes, with a substantial increase in *Affective Boost High* and *Elaborated Schema* nodes, and a significant corresponding decrease in *Affective Boost Low* and *Inert Schema* nodes. In contrast, the *Experiment 3 DI model* (Figure 8) still shows no increase in *Affective boost*, but exhibits a greater decrease in the *No Assembly* node and a minimal increase in *Elaborated Schema*, while all other nodes remain unchanged. These findings suggest that the benefits of Productive Failure become increasingly pronounced as instructional fidelity improves—particularly in supporting the transfer of learning to novel contexts, an area where Direct Instruction consistently underperformed.

The results of these computational experiments provide preliminary validation for aspects of the model that align with empirical findings—particularly the pattern of PS-I outperforming I-PS—consistent with the direction of results reported by Sinha and Kapur (2021). In Experiment 2, the knowledge gain for Productive Failure (0.49) was approximately 9.8 times greater than that for Direct Instruction (0.05), which exceeds the relative advantage observed in Sinha and Kapur's meta-analysis, where PF was found to be 1.8 times more effective in high-fidelity implementations. In Experiment 3, PF outperformed DI by a factor of 2.9 in knowledge acquisition (0.73 vs. 0.25) and demonstrated clear superiority in transfer performance (0.59 vs. 0.00). These results align with the direction of empirical findings, although the transfer advantage in the computational model appears even more pronounced—possibly reflecting an amplification of PF's benefits under idealized implementation conditions. Such discrepancies between modeled and empirical outcomes also serve as valuable opportunities for iterative model refinement. For example, if the model overestimates transfer gains under Productive Failure, this may indicate the need to review and adjust how cognitive processes like schema elaboration or motivational factors such as affective boost are parameterized. By comparing simulated outcomes with real-world data, the model can be progressively refined, reflecting a cyclical, adaptive, and evolutionary approach to computational educational research. This iterative process can help enhance the model's validity and support its use as a predictive and explanatory tool for future instructional design research.

Conclusion

This paper reports on research that applies computational modeling as a tool in educational research. By modeling classroom learning as a complex system, this approach enables the exploration of how individual-level factors—such as affective boost, prior knowledge, and schema development—interact to produce learning and transfer outcomes. We began by validating our model using real-world empirical data, then conducted computational experiments to explore new variable configurations beyond those yet tested in empirical studies. This process illustrates how computational modeling can inform the design of future research, guiding the development of research questions, predicted outcomes, and experimental structures across diverse educational contexts. We advocate for broader adoption of computational methods as a complement to existing quantitative and qualitative approaches in educational research. Such integration may deepen our understanding of dynamic learning processes and contribute to evidence-based decision-making in education.

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition* (2nd ed.). National Academies Press. https://doi.org/10.17226/9853

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152. https://www.sciencedirect.com/science/article/pii/S0364021381800298

Denzin, N. K., & Lincoln, Y. S. (2018). *The discipline and practice of qualitative research*. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 1–26). SAGE Publications.

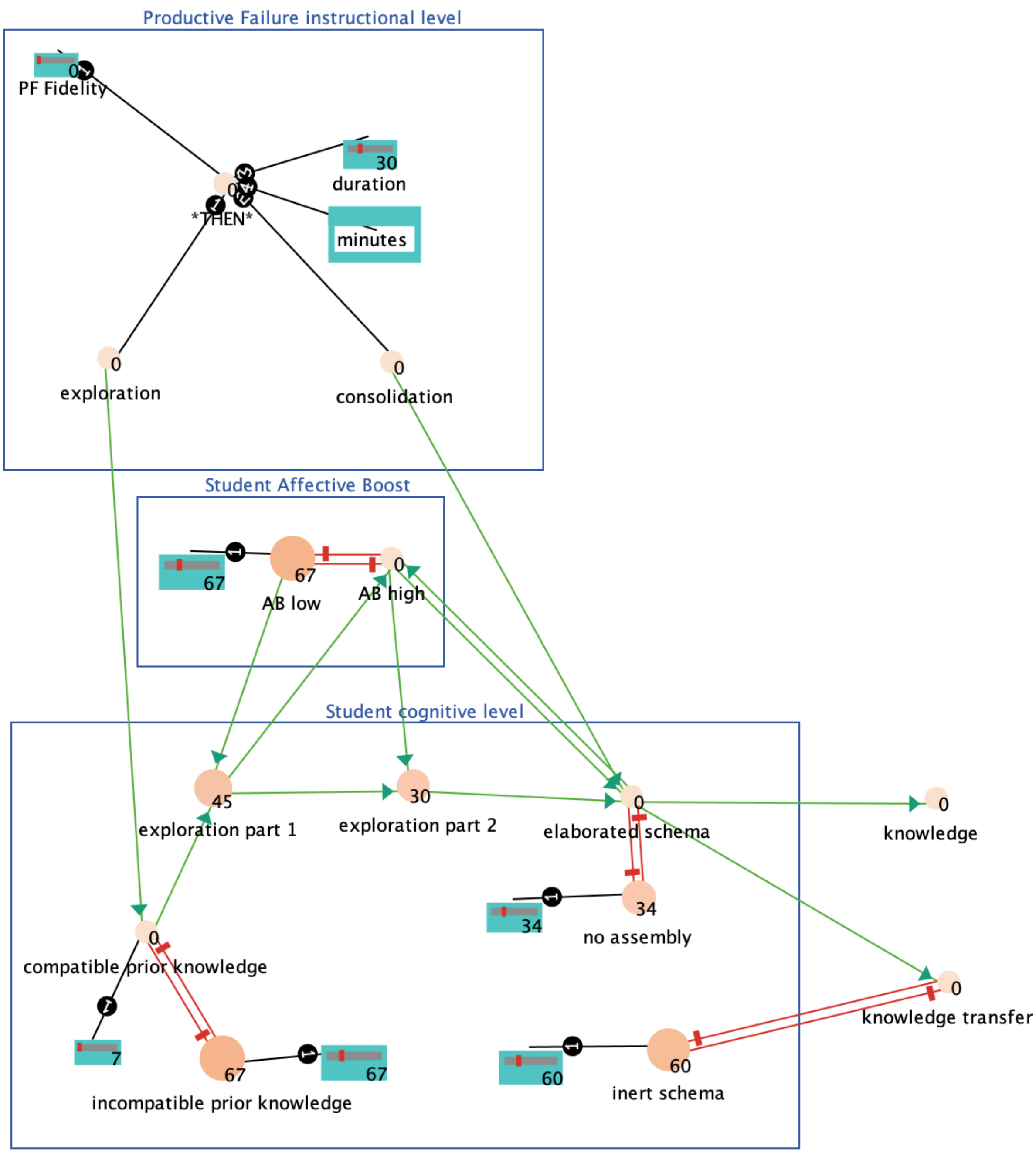
Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. Irvington Publishers.
Hedges, L. V. (1981). Distribution theory for Glass's Estimator of Effect Size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.2307/1164588
Jacobson, M. J. (2020). Complexity conceptual perspectives for research about educational complex systems. *The Journal of Experimental Education, 88*(3), 375-381. https://doi.org/10.1080/00220973.2019.1652138
Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing and investigating learning in complex and open-ended learning environments. *Educational Technology Research and Development, 64*(2), 285–306. https://doi.org/10.1007/s11423-015-9411-0
Jacobson, M. J., Levin, J. A., & Kapur, M. (2019). Education as a complex system: Conceptual and methodological implications. *Educational Researcher, 48*(2), 112-119. http://doi.org/10.3102/0013189x19826958
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. http://doi.org/10.1016/j.learninstruc.2017.03.007
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26. https://doi.org/10.3102/0013189X033007014
Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669
Kapur, M. (2024). *Productive Failure: Unlocking deeper learning through the science of failing* (1st ed.). Wiley. https://doi.org/10.1002/9781394308712
Levin, J. A., & Datnow, A. (2012). The principal role in data driven decision making: Using case study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement, 23*(2), 179-201. https://doi.org/10.1080/09243453.2011.599394
Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3
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
Smith, P. (2007). *Systems thinking and systems dynamics.* Emerald Group Publishing.
Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling. Northwestern University. http://ccl.northwestern.edu/netlogo.
Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo.* The MIT Press. https://mitpress.mit.edu/9780262731898/an-introduction-to-agent-based-modeling/
Williamson, B., Potter, J., & Eynon, R. (2019). New research problems and agendas in learning, media and technology: The editors' wishlist. *Learning, Media and Technology*, *44*(2), 87-91. https://doi.org/10.1080/17439884.2019.1614953

Figures and Tables

Table 1
Independent Variable Settings and Dependent Variable Results for Three Computer Experiments. Instructional fidelity effect sizes from Sinha & Kapur (2021).

Experiment	**Independent Variables**		**Dependent Variables**		
				PF	**DI**
Pretest	**Instructional Fidelity Effect Size**	**0.00**			
	AB Low	0.67	**Knowledge**	**0.00**	**0.00**
	Compatible Prior Knowledge	0.07			
	Incompatible Prior Knowledge	0.67			
	Inert Schema	0.60	**Transfer**	**0.00**	**0.00**
	No Assembly	0.34			
1	**Instructional Fidelity Effect Size**	**0.20**			
	AB Low	0.67	**Knowledge**	**0.08**	**0.00**
	Compatible Prior Knowledge	0.07			
	Incompatible Prior Knowledge	0.67			
	Inert Schema	0.60	**Transfer**	**0.00**	**0.00**
	No Assembly	0.34			
2	**Instructional Fidelity Effect Size**	**0.36**			
	AB Low	0.67	**Knowledge**	**0.49**	**0.05**
	Compatible Prior Knowledge	0.07			
	Incompatible Prior Knowledge	0.67			
	Inert Schema	0.60	**Transfer**	**0.16**	**0.00**
	No Assembly	0.34			
3	**Instructional Fidelity Effect Size**	**0.58**			
	AB Low	0.67	**Knowledge**	**0.73**	**0.25**
	Compatible Prior Knowledge	0.07			
	Incompatible Prior Knowledge	0.67			
	Inert Schema	0.60	**Transfer**	**0.59**	**0.00**
	No Assembly	0.34			

Figure 1
Screenshot of the PF Model link/node states for the pretest, with instructional fidelity effect size set to 0.00.



PF pretest

Figure 2

Screenshot of the DI Model link/node states for the pretest, with instructional fidelity effect size set to 0.00.

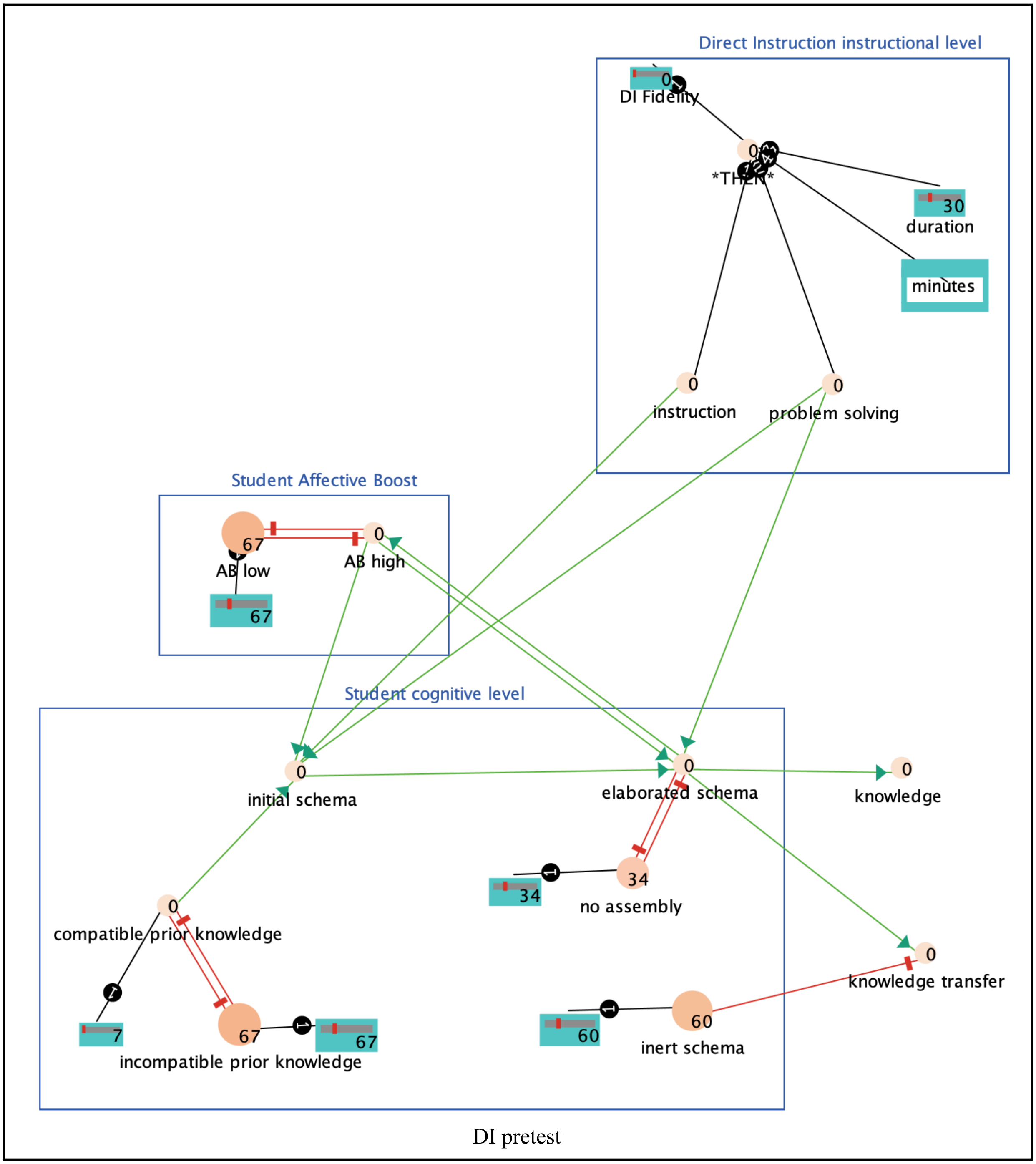


Figure 3

Screenshot of the PF Model link/node states for Experiment 1, with instructional fidelity effect size set to 0.20.

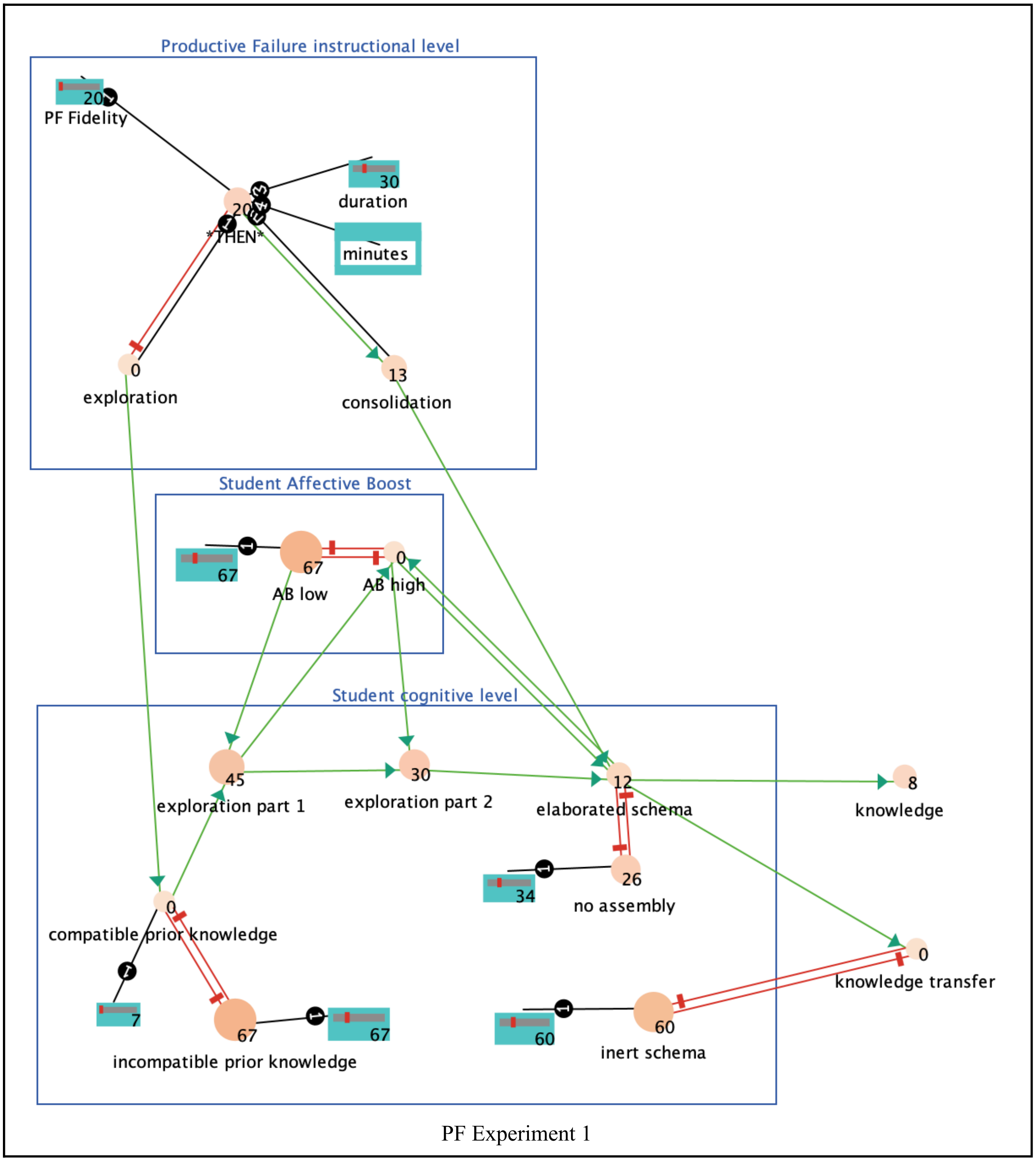


Figure 4
Screenshot of the DI Model link/node states for Experiment 1, with instructional fidelity effect size set to 0.20.

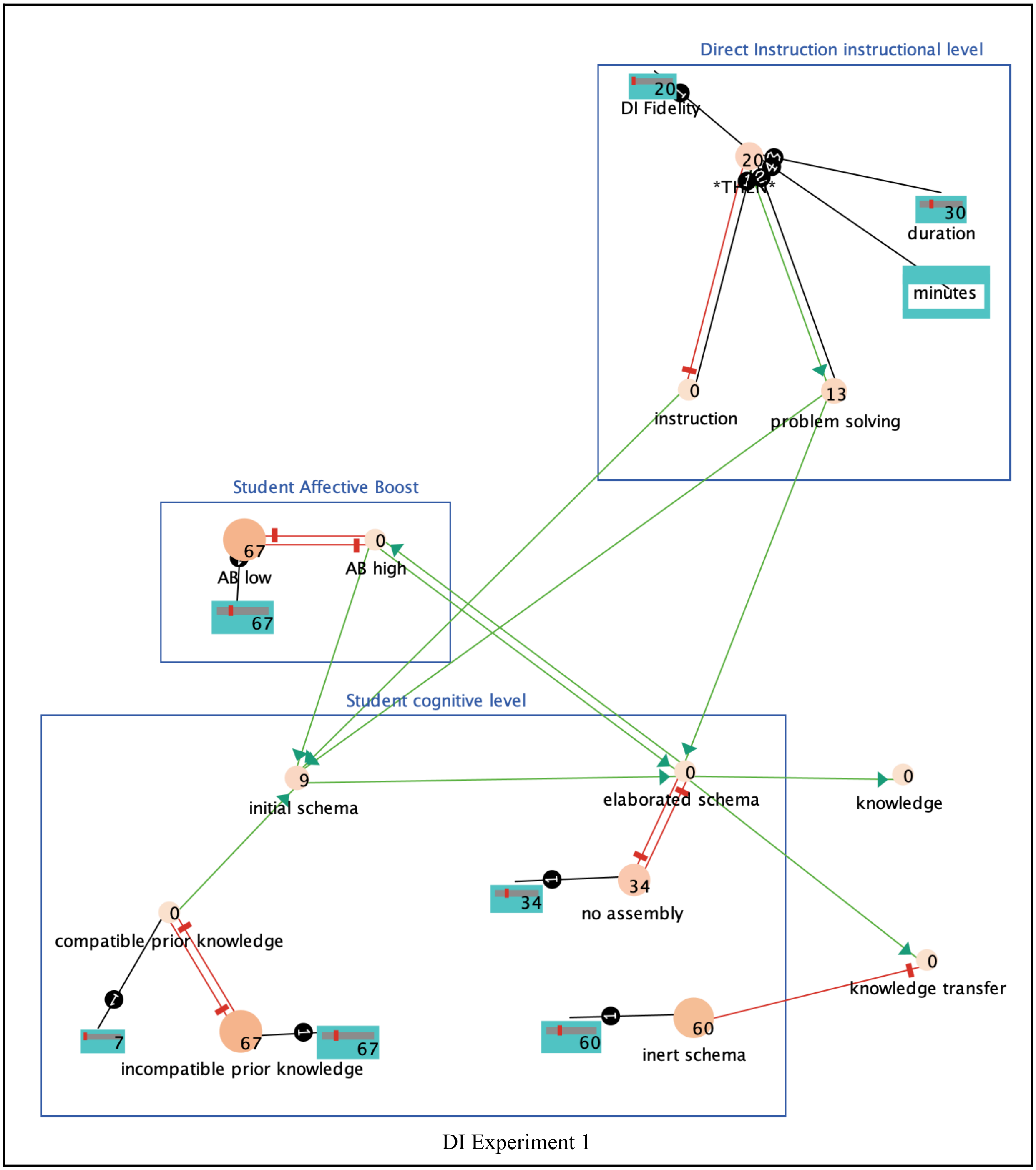


Figure 5

Screenshot of the PF Model link/node states for Experiment 2, with instructional fidelity effect size set to 0.36.

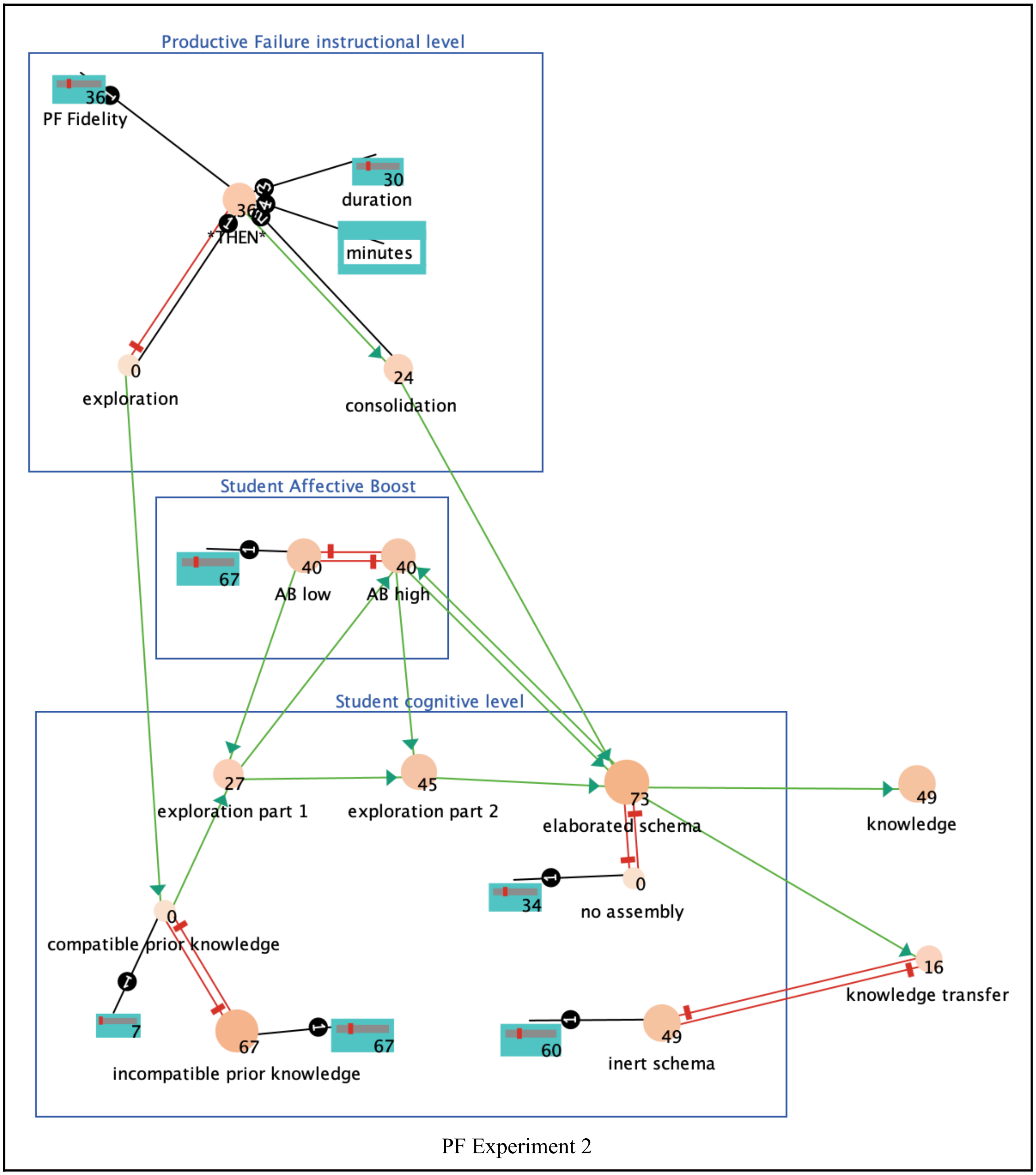
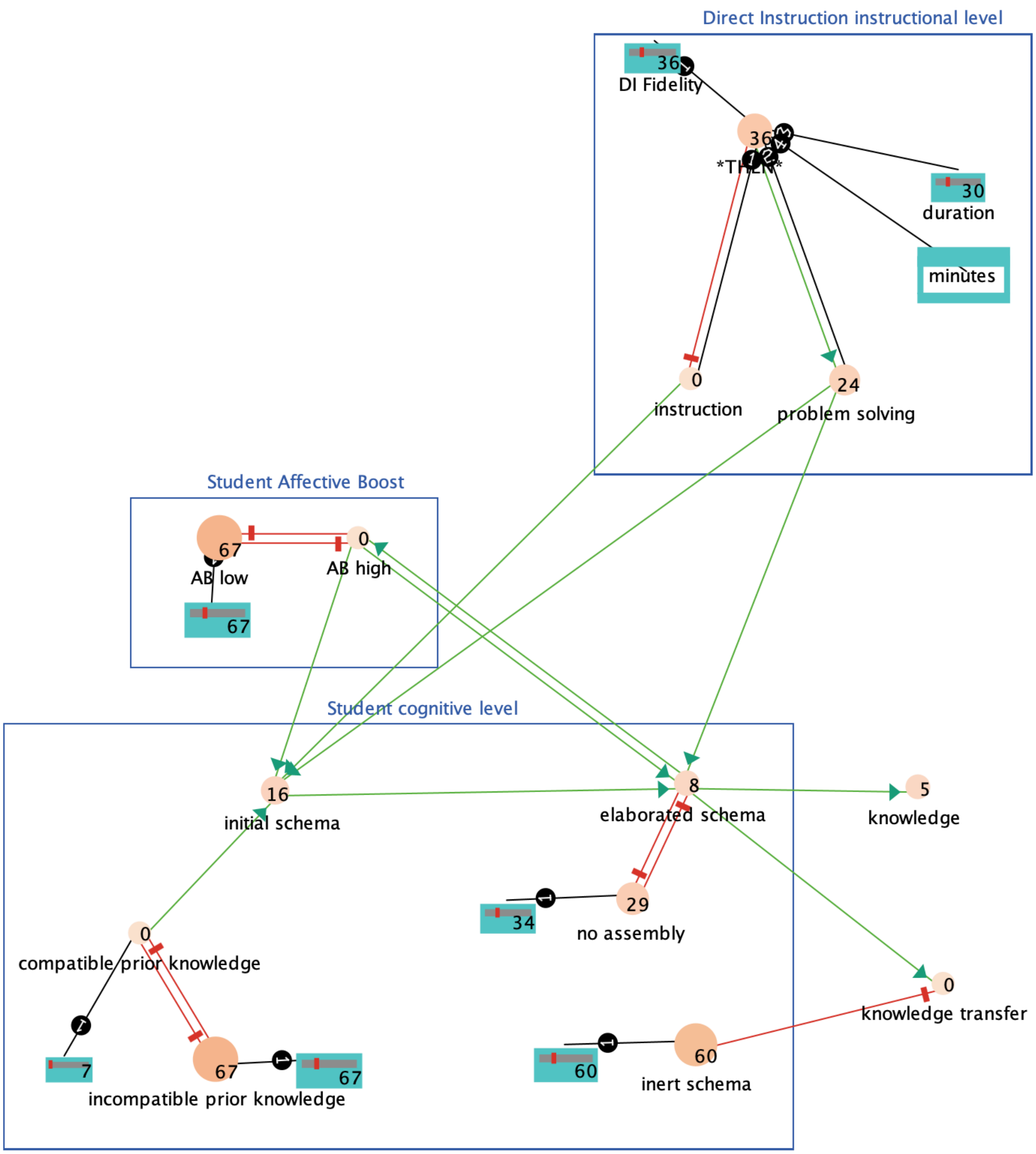


Figure 6

Screenshot of the DI Model link/node states for Experiment 2, with instructional fidelity effect size set to 0.36.



DI Experiment 2

Figure 7

Screenshot of the PF Model link/node states for Experiment 3, with instructional fidelity effect size set to 0.58.

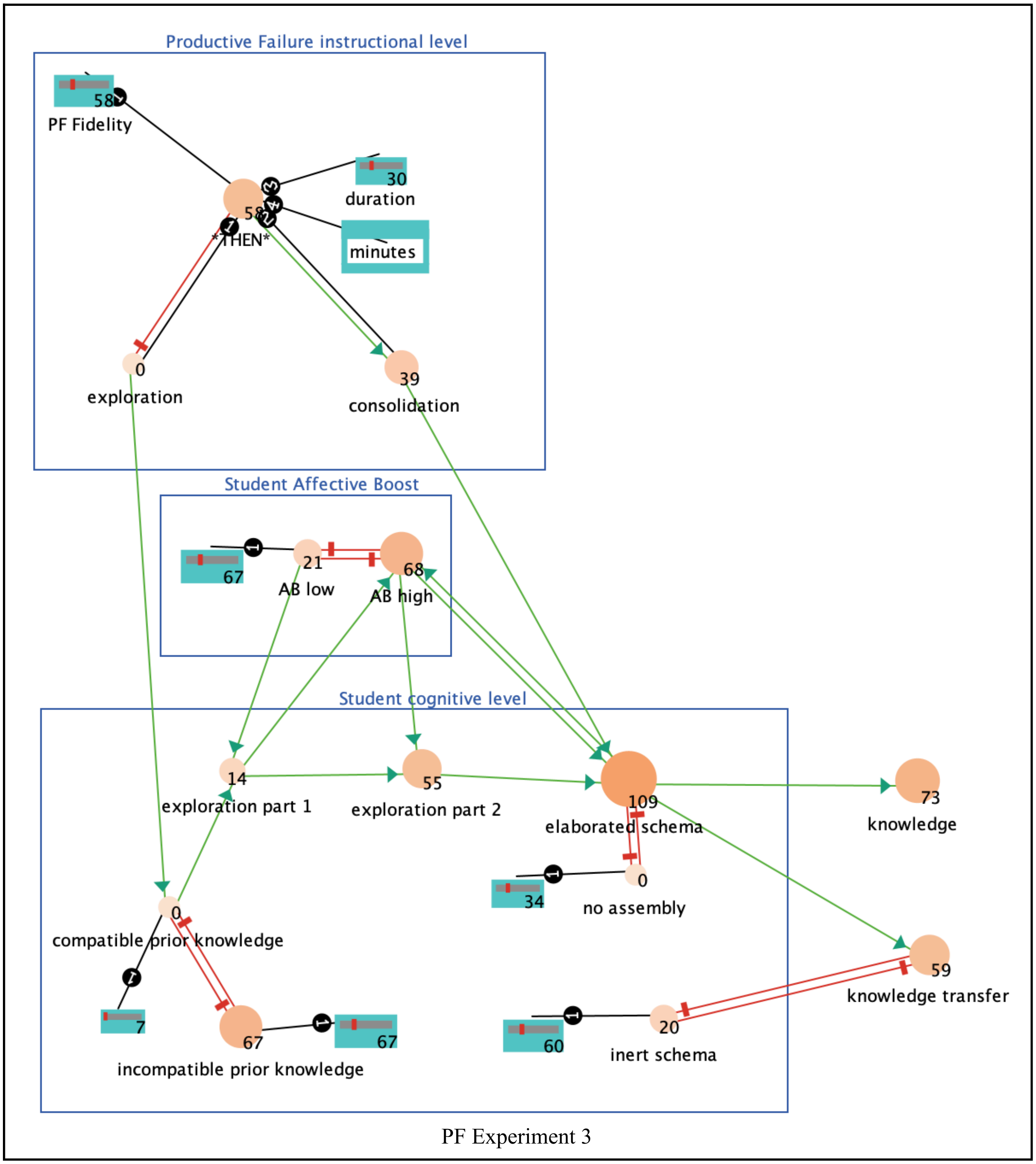
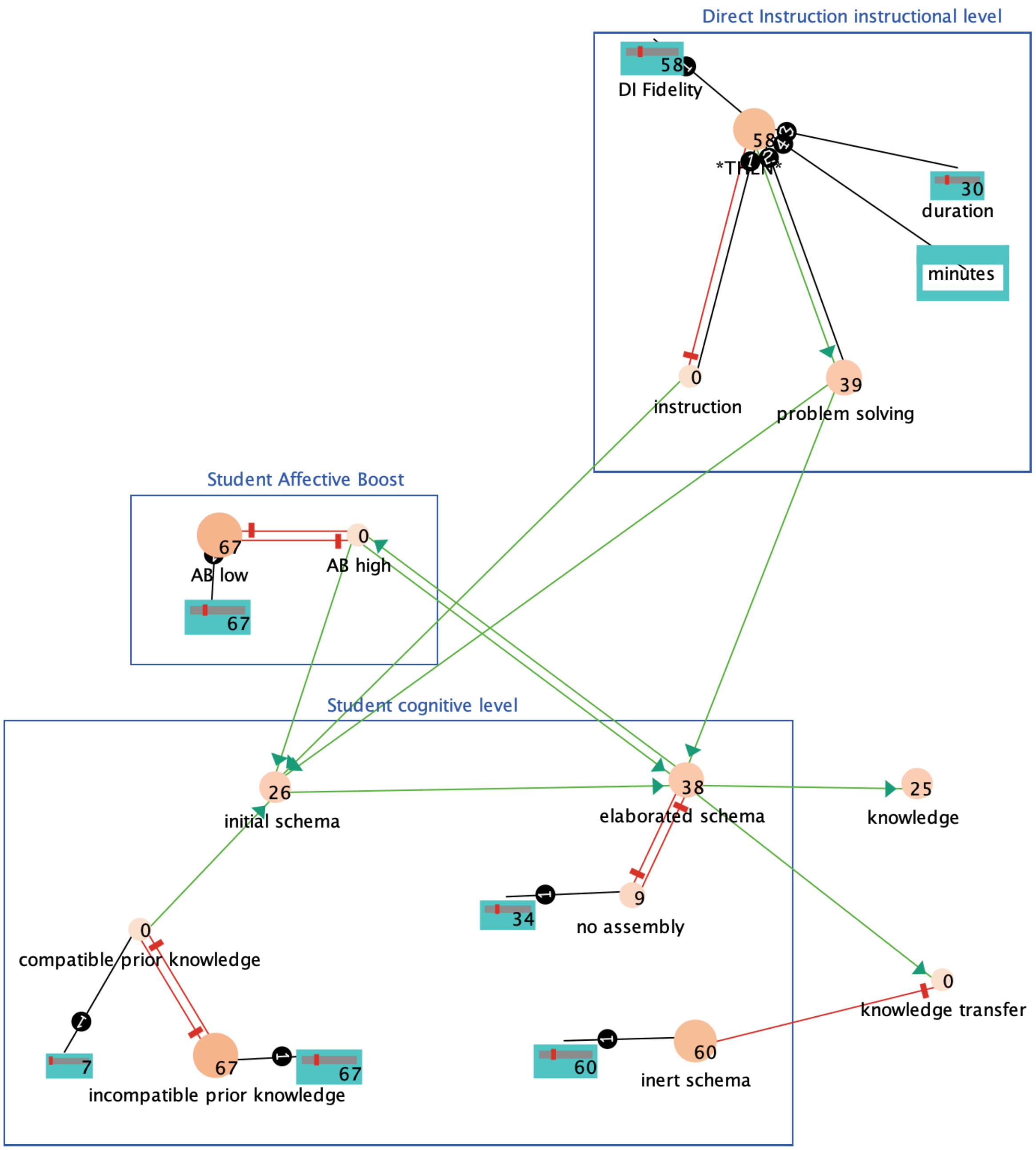


Figure 8

Screenshot of the DI Model link/node states for Experiment 3, with instructional fidelity effect size set to 0.58.



DI Experiment 3