

Is it Better to Give than to Receive? The Assistance Dilemma as a Fundamental Unsolved Problem in the Cognitive Science of Learning and Instruction

Kenneth R. Koedinger, Phillip Pavlik, Bruce M. McLaren, Vincent Alevan

(koedinger@cmu.edu, ppavlik@cs.cmu.edu, bmclaren@cs.cmu.edu, alevan@cs.cmu.edu)

Human-Computer Interaction Institute
 5000 Forbes Avenue
 Carnegie Mellon University
 Pittsburgh, PA, 15213-3891 United States

Abstract

To foster more robust student learning, when should instruction provide information and assistance to students and when should it request students to generate information, ideas, and solutions? In different forms, this dilemma for instructors has been a part of debates on education since Plato. However, it is fair to say that we remain far from a precise and sound scientific response. We believe this “Assistance Dilemma” is one of the fundamental unsolved problems in the cognitive and learning sciences. To address this dilemma, we suggest a four step strategy for more clearly articulating the problem and tackling it with computational models that can be used to make precise, replicable, and testable predictions about when instructional assistance should be given vs. withheld. We illustrate these steps on two different dimensions of instructional assistance. On the “problem spacing” dimension, we present a computational model that generates precise predictions of the kind we call for. On the more complex “example-problem” dimension, we illustrate how the field is at a point where such a precise computational model may be possible.

Keywords: Learning; problem solving; cognitive modeling; computational modeling, cognitive psychology; education.

The Assistance Dilemma as an Open Problem

This paper discusses the Assistance Dilemma: “How should learning environments balance information or assistance *giving* and *withholding* to achieve optimal student learning?” (Koedinger & Alevan, 2007). This question presents a dilemma not only because numerous experimental results sometimes indicate benefits and other times indicate costs of greater instructional assistance, but also because we lack sufficient cognitive theory to predict when instructional assistance will be beneficial or harmful.

The goal of resolving the Assistance Dilemma is to have a predictive theory of what instructional methods best achieve “robust learning”. Robust learning is operationalized by one or more post-instruction measures: transfer, long-term retention, or accelerated future learning. Because time is so valuable for students and instructors, we are also concerned with learning efficiency, that is, how much instructional time is needed to achieve robust learning outcomes.

We describe the Assistance Dilemma as a fundamental open research problem for the learning sciences. We define assistance broadly to not only include explicit instructional guidance or scaffolds, but also any change in the

instructional environment that increases immediate performance or reduces mental effort. Thus, a change that may put greater demands on the learner (a difficulty) during instruction is lowering assistance.

Table 1 illustrates the Assistance Dilemma by highlighting how the level of assistance during instruction is not correlated with learning outcomes. As shown in the first row, sometimes instructional assistance can be a “crutch” that harms learning (e.g., if I always tie my child’s shoes, she will never learn how to do it on her own) while sometimes assistance can be a “scaffold” that bootstraps learning (e.g., if I show my child how to tie her shoes, she will have an example from which to learn herself). Notice that our use of “assistance” describes methods and affordances employed during instruction. To reduce potential confusion, we will not use “assisting” learning but “improving” or “enhancing” learning when referring to the longer-term consequences of instruction on future student performance outside the instructional environment.

Table 1: Assisting Performance During Instruction May Aid or Harm Learning

Instructional support	Poor learning outcome	Better learning outcome
High assistance (less demanding)	crutch	scaffold
Low assistance (more demanding)	undesirable difficulty; extraneous load	desirable difficulty; germane load

The second row of Table 1 illustrates how lower levels of assistance (or inversely putting greater demands on students during instruction) can sometimes lead to poorer learning and other times lead to better learning. A long line of research on “cognitive load theory” (e.g., Sweller, Van Merriënboer, & Paas, 1998) suggests how some typical forms of instruction, like homework practice problems, put “extraneous” processing demands (or “extraneous load”) on students that may detract from learning. Higher levels of assistance or guidance, for instance in the form of more frequent use of worked solution examples, lead to both more efficient learning and better transfer.

However, another line of research on “desirable difficulties” suggests ways in which making task performance harder during instruction, for instance, by delaying feedback, enhances learning (Schmidt and Bjork, 1992). And even within the cognitive load line, researchers have identified situations where lower assistance or greater demands lead to be better learning. For instance, Paas and Van Merriënboer (1994) found that while worked examples made performance easier during instruction *and* led to better learning, introducing greater variability in example content made performance *harder* during instruction but also led to better learning. The researchers suggested that some instructional forms reduce extraneous cognitive load, like worked examples in this case, but others, like example variability increase “germane” cognitive load.

The Assistance Dilemma is at the heart of many “education wars” with “traditionalists” typically advocating forms of greater assistance (e.g., Kirshner, Sweller, Clark, 2006; Mayer, 2004) and “reformers” advocating less assistance (e.g., Jonassen, 1991). Long-standing notions like zone of proximal development (Vygotsky, 1978), aptitude-treatment interactions (Cronbach & Snow, 1977), or model-scaffold-fade (Collins, Brown, & Newman, 1990) suggest that instructional assistance should be greater for beginning learners and be reduced as student competence increases.

Such long-standing notions and the more recent extensive experimental research on cognitive load, desirable difficulties, etc. might lead one to wonder “what’s the dilemma?” Just give novices more assistance and fade it away as they become more expert. However, current theory does not predict *how much* assistance to initially provide nor *when* and *how fast* to fade it. Further, it does not provide predictive guidance as to *when* an instructional demand is “germane” or “extraneous”, “desirable” or “undesirable.” Despite much relevant research and many different characterizations of the fundamental problem, the Assistance Dilemma remains unresolved because we do not have adequate cognitive theory to make *a priori* predictions about what forms and levels of assistance yield robust learning under what conditions. We need to get beyond over-simplified dichotomies and “tighten the inferential web that ties experimental studies together.” (Newell, 1973)

A Plan of Attack

Defining Assistance Levels

Higher assistance during instruction may come in the more direct or explicit form of providing more information to students, for instance, by showing or telling them an idea rather than asking them to generate it themselves. However, in our conceptualization, higher assistance can also come implicitly from any kind of instructional affordance or scaffold that makes an instructional activity easier for students. The higher the assistance, the less mental effort required on the part of the student and/or the more likely the student performs correctly on the *immediate* instructional activity. Of course, the key point of the Assistance Dilemma

is that just because higher assistance makes local performance better (or reduces mental effort) does not mean it will lead to enhanced learning. An interesting illustration of this point comes from the human-computer interaction literature where Gilmore (1996) demonstrated that changes to a computer interface that improve user performance do not necessarily improve learning. For instance, a more direct manipulation interface for controlling a transportation system simulation made control easier, but participants learned less about the system than those using a more cumbersome interface.

Steps Toward Resolving the Assistance Dilemma

As indicated in the examples above, there are many possible dimensions of assistance including giving lots of example solutions vs. withholding them (problems), giving vs. withholding immediate feedback, giving low vs. high variability examples. The first step in our suggested plan of attack for addressing the Assistance Dilemma is to select *a single* dimension of assistance on which to focus an extensive research program. We are *not* going to find a generic resolution that works across all dimensions.

The second step is to collect, summarize, and integrate the relevant literature on that dimension. What studies have explored different levels of assistance and what have the effects been on measures of robust learning and learning efficiency? Such integration has been a key goal of the Pittsburgh Science of Learning Center and we have been using a wiki (learnlab.org/research/wiki/) for a broader community to share in this process of collection, summarization, and integration.

The third step is to characterize a set of conditions and parameters that can be used as part of a precise theoretical model that makes “computable” predictions about robust learning efficiency. Whether such a model is a mathematical model or a full cognitive model embedded in a cognitive architecture is perhaps less important than its being usable by other researchers to make precise and replicable predictions. This step is by no means an easy one and, in fact, the main goal of this paper is to frame the challenge and illustrate it with two examples of efforts at different stages of progress

Of course, a theoretical model should not only be able to explain and integrate past results, but make new predictions. Thus, the fourth step is to apply and test the model in new settings, domains, and/or with new student populations. The model should be used to make *a priori* predictions (i.e., before a study is run) about the level of assistance (fixed or adaptive) that will achieve optimal robust learning efficiency. Then new experiments should be run to test those predictions.

A Success Case: The Practice Spacing Dimension of Instructional Assistance

We first illustrate our plan of attack on the “practice spacing” dimension of instructional assistance. Practice spacing is the time interval between successive practices of

a to-be-learned fact. For instance, if I'm learning Chinese vocabulary by studying Chinese-English word pairs, the practice spacing is the time between when I practice a particular pair (e.g., xie2zi to shoes) and when I practice that pair again. A straightforward instructional prescription based on the "spacing effect" is that practice should be as widely spaced as possible (Pashler et al., 2007). By this view, wide spacing is a kind of "desirable difficulty"—while it makes successful retrieval more difficult during instructional practice, it leads to more robust learning, particularly long-term retention. This conclusion is supported by a great amount of research (e.g. Pashler, Zarow, & Triplett, 2003). However, this research usually fails to note that the higher level of difficulty during spaced practice results in more errors and an increased need to consume valuable practice time in error correction and reteaching (Pavlik Jr. & Anderson, accepted). This downside of wide spacing is particularly apparent early in learning when the probability of error after a long retention interval is quite high.

Thus, we have an instance of the Assistance Dilemma: Should instruction provide more assistance, in the form of shorter practice intervals, so as to make learning trials proceed more quickly or should instruction provide less assistance, in the form of longer spacing intervals, so as to maximize the gain in long-term retention for each trial? Following our plan of attack what we need is a set of conditions and parameters that provide for theoretical predictions about how much robust learning is achieved per instructional unit for different levels of assistance on this dimension.

Pavlik and Anderson (accepted) has done just that by developing a model that characterizes these conditions and parameters for a paired-associate practice task (e.g., learning Chinese-English word pairs) similar to tasks used by other researchers of spacing effects (e.g. Pashler, Cepeda, Wixted, & Rohrer, 2005). Because the model is quantitative (based on the ACT-R architecture, Anderson & Lebiere, 1998; Pavlik Jr. & Anderson, 2005), it can precisely characterize the conditions of spacing, recency from prior practice, frequency of practice, and history of prior success to make predictions about the future effect of more practice as a function of these conditions. Having this quantitative model that relates current practice to future performance (i.e., learning) allows one to compute directly the best resolution to the assistance dilemma in the form of the optimal spacing interval.

The key to this method is the realization that it should account for instructional time or "time on task" and characterize the cost-benefit trade-offs that best predict efficiency in robust learning. The idea that cost-benefit information must be considered is not new (Atkinson & Paulson, 1972), but Pavlik's fine-grained formalization of the cost-benefit structure for spaced practice is a unique example of a method that allows us to resolve the assistance dilemma in a specific domain. Equation 1 predicts robust learning efficiency gain (eff_m) to be achieved for studying a

particular fact at a particular time based on a history of prior practice opportunities. We are interested in how different levels of assistance, in this case the length of the spacing interval, affect this equation. When all else is equal (e.g., frequency of prior practice), the spacing interval is monotonically related to the "activation level" of a fact in memory (the m in Equation 1): The longer the spacing, the lower activation (the learner forgets). Thus, the x-axis in Figure 1 can be read as either the level of assistance (high to low), the spacing interval (narrow to wide), or the activation level of a fact after that spacing level has been experienced.

$$eff_m = \frac{p_m b_{suc} g_m + (1 - p_m) b_{fail} g_m}{p_m (t_m + fixedsuccosts) + (1 - p_m) fixedfailcosts} \quad (1)$$

In Equation 1 p_m is the probability of recall, t_m is the time cost of recall, b_{suc} scales the gain from a successful recall, b_{fail} scales the gain from the review after a failure to recall, g_m captures the long-term increase in activation at a specific desired retention interval, $fixedsuccost$ captures the perceptual motor costs of practice when the result is success and $fixedfailcost$ captures the time costs of failure and the review following failure. The activation is the m in the subscripts of key variables p , b , and g indicating they are all a function of m . The Equation can be read as the expected learning benefit of retrieval success ($p_m * b_{suc} * g_m$) or failure ($(1 - p_m) * b_{fail} * g_m$) divided by the expected instructional time costs of success ($p_m (t_m + fixedsuccosts)$) or failure ($(1 - p_m) fixedfailcosts$).

To determine whether an item being learned is at the point of maximal future learning per second of current practice, Equation 1 is optimized. The behavior of this equation depends only on the activation of the item being learned. If activation is high (which the model says occurs when there have been many practices, recent practice, previous widely spaced practices, or more successful practice) then p_m is high, t_m is low, and g_m is low. Because we have this equation we can compute Figure 1, which shows how the efficiency computation (eff_m) depends on activation. The shape of the function depends on model parameters that are determined by fitting the model to prior student data.

As Figure 1 shows there is an optimum point at which learning efficiency is maximal, that is, learning per second is optimized. According to this figure, facts should receive practice whenever their activation is at the optimal point. Equation 1 (and the other equations behind it) represent a major step (step #3 above) toward the resolution of the assistance dilemma as applied to spaced practice of simple facts. Having this computational model also allows the researcher to run simulation studies and make predictions. In the case where the experiment being simulated is very similar to old experiments such predictions can be fairly accurate and have been shown to capture at least the general patterns of later data gathered in the experiment simulated (Pavlik Jr. & Anderson, accepted). In this experiment Pavlik and Anderson began with a model of task conditions, student variability, and specific task parameters and used

this information to simulate students running through the experimental conditions (e.g., an optimized condition and a wide spacing condition). The simulated student model was given experimental tasks and generated either correct or incorrect answers as predicted by the model. The simulated student also produced latencies that allowed a simulation of time on task. The optimized condition contained a separate student-tracking model that uses the Equation 1 to pick the next fact for the student to practice. This student-tracking model picks a fact, which based on the time since its last practice, is closest to the maximum gain point shown in figure 1.

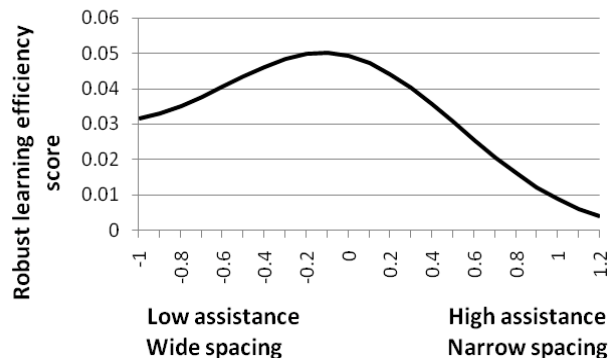


Figure 1. Robust learning efficiency score as a function of the level of assistance on the “practice spacing” dimension. The x-axis varies from low assistance, which is achieved by wide spacing that results in low activation of a fact at the time of instruction, to high assistance, which is achieved by narrow spacing that results in higher activation.

Such predictive simulation is not limited to situations where all the prior parameters have been determined experimentally. If good parameter guesses for another task can be made, the simulation can be used to estimate the shape of Figure 1 in that task. For example, if the failure costs are very low (e.g., as in training recognition rather than recall), the inverted-u shape disappears and wide spacing is optimal. Similarly, if forgetting is negligible (e.g., because highly-integrated knowledge structures can be quickly acquired), then the model no longer captures a benefit for spacing and massed practice is predicted as optimal. These examples highlight how the model can be used to make novel predictions and that the approach may not always recommend a middle ground solution.

This approach has been implemented in a computerized training program (<http://optimalllearning.org>) and studies of its use in practice are showing learning benefits, for instance, in learning Chinese vocabulary (Pavlik Jr., Bolster, Wu, Koedinger, & MacWhinney, 2008).

A Case in Progress: The Example-Problem Dimension of Instructional Assistance

Another assistance dimension that we are exploring is the example-problem dimension. This dimension involves a

continuum in which students are provided with problems, examples, or combinations of problems and examples. From lowest to highest assistance, points along this dimension include (a) problem solving only, (b) integrated worked examples and problems, (c) examples only.

Many prior experiments (e.g., Paas, 1992; Schwonke et al, 2007; Sweller & Cooper, 1985; Trafton & Reiser, 1993) point to a learning advantage for greater use of worked examples (b or c beat a) with advantages typically shown in both learning efficiency and transfer. Summarizing this work, Clark & Mayer (2003) articulated the worked-example principle: “Replace some practice problems with worked examples.” The practical power of this principle has been particularly highlighted in a series of recent lab and classroom *in vivo* experiments. Greater use of examples in already well-proven intelligent tutoring systems has been experimentally shown to lead to even further gains in robust learning efficiency in studies of student use of intelligent tutors in geometry (Schwonke et al, 2007), physics (Ringenberg & VanLehn, 2006), and chemistry (McLaren, Lim, & Koedinger, 2008).

Cognitive load theory is the leading explanation of the worked example principle (c.f. Sweller, Van Merriënboer, & Paas, 1998). This theory suggests that problem solving produces “extraneous cognitive load” because of the resources needed to store and manage problem-solving goals. With their working memory so stressed, students have little left to engage in generative learning processes. In contrast to problems, worked examples free those resources for learning processes like the induction and refinement of new knowledge components.

Cognitive load theory provides a clear explanation for the benefits of examples, but does explain why including problems in instruction should be beneficial. It seems to suggest that worked examples alone, a high-assistance approach, would be the least taxing on cognitive resources and thus best for learning. In fact, few studies have made the relevant direct comparison of examples only (c above) vs. integrated examples and problems (b). We only know of one such study (Stark, Gruber, Renkl, & Mandl, 2000) and it demonstrated that, contrary to a straightforward application of cognitive load theory, learning from the examples-only condition was significantly worse than learning from the integrated examples and problems condition. Contrary to calls for maximal guidance (Kirshner et al., 2006), it appears that *mid-level* assistance (integrated examples and problems) leads to better robust learning efficiency than lower (all problems) or higher (all examples) levels of assistance. There are also strong theoretical reasons for why problems have some, perhaps complementary, benefits over examples including the generation and testing effects (e.g., Hausmann & VanLehn, 2007; Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel, et al., 2007).

Like the levels on the practice spacing dimension, the accumulating data on points along the example-problem dimension provide a starting point for developing a predictive theory. Unlike the practice spacing dimension, we

are only beginning to formulate parameters and mechanisms that would lead to an equation like Equation 1 (see more below). Without such a computational model, we do not have a reliable and replicable way to make predictions about the benefits and costs of different combinations of examples and problems for students at different levels of competence.

Nevertheless, the evidence thus far seems to indicate that such an equation would be plotted as an inverted-U such as the abstract graph shown in Figure 2. It appears the greatest robust learning efficiency is achieved in the mid-level with a balanced use of examples and problems and learning is either or both worse and less efficient when the assistance is higher (all examples) or lower (all problems).

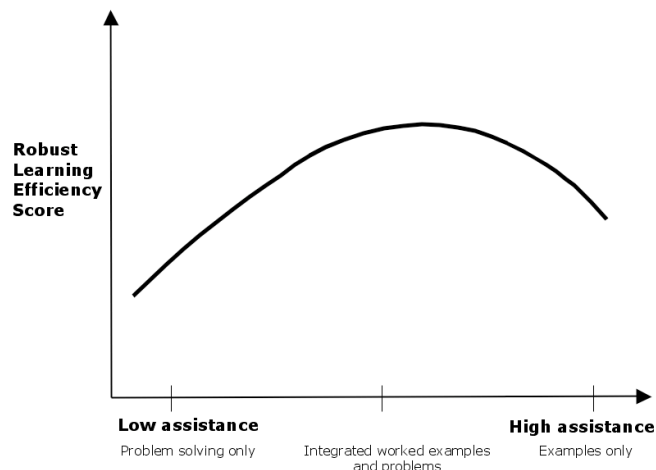


Figure 2. Abstract graph of the example-problem dimension based on existing evidence.

With respect to our “plan of attack” on this dimension, we are engaging in step 3, attempting to clearly define the parameters and conditions that are relevant to formulating a predictive theory. In our efforts it has become clear that we need more experiments that explore a range of possible combinations and orderings of examples and problems. To clearly establish the inverted-U, we need all three conditions in a single study: all problems, integrated examples and problems, and all examples. To our knowledge, such a three-condition “bracketing” study has not been performed.

Nevertheless, the large body of experimental evidence collected within this dimension highly suggests an inverted-U curve such as the abstract graph shown in Figure 2. While we have yet to formulate a variable (or set of variables) comparable to “activation” on the practice spacing dimension, the activation and/or utility metrics in ACT-R are reasonable candidates. In early stages of learning, ACT-R predicts that students learn by solving problems by analogy whereby they recall or seek out similar examples and use them as a basis for analogical inference (Anderson, Fincham, & Douglass, 1997). By studying isomorphic examples first, as in the mid-level of Figure 2, students may be increasing the activation of relevant components of these examples that can then be better recalled during analogical problem solving. In a

complementary way, subsequent problem solving provides opportunities for learners to prune (or lower the utility of) incorrect knowledge that inevitably results from the inductive character of analogical processes (c.f., Matsuda, Cohen, & Koedinger, 2008).

This analysis is just a hint of the direction that needs to be pursued to create a predictive model for this dimension. Building on Schwonke et al (2007), on-going work is exploring the benefits of adaptive “fading” of examples (given steps) into problems (requested steps). Such work not only requires the creation of a candidate computational model, but is also providing fine-grain longitudinal log data of individual learner interactions. These provide the basis for more precise characterization of conditions and parameters for predictive modeling.

Other Assistance Dimensions

To fully address the Assistance Dilemma, the many other dimensions of instructional assistance must be examined through extended programs of experimental research, integrative reviews, and the development of precise computational models. Examples of other assistance dimensions include feedback timing, example variability, interface affordances, implicit and explicit instruction, scripted and unscripted peer collaboration, self-explanation prompting, metacognitive scaffolds, etc. The Pittsburgh Science of Learning Center is supporting researchers from Germany to California in developing a better understanding of robust learning and exploring the Assistance Dilemma along a number of such dimensions (see learnlab.org/research/wiki).

Conclusions

The Assistance Dilemma raises the question of whether in instruction it is better “to give or to receive.” It is not a new idea *per se*, but we think it is critical to a) recognize it as a fundamental unsolved problem in cognitive science and b) frame a strategy for addressing this problem. We presented a four-step strategy and illustrated two cases of pursuing the strategy. The analysis presented in this paper raises serious questions about framing debates about instruction as a binary choice (e.g., direct instruction vs. constructivism). While such dichotomizing engenders attention and energy, it does not advance scientific understanding nor the development of productive methods for instructional engineering.

Acknowledgments

This research and theoretical integration was supported in part by the Pittsburgh Science of Learning Center, NSF Grant # 0354420.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: LEA..

- Atkinson, R. C., & Paulson, J. A. (1972). An approach to the psychology of instruction. *Psychological Bulletin*, 78(1), 49-61.
- Collins, A., Brown, J.S., & Newman, S.E. (1990). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. & Snow, R. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington Publishers.
- Gilmore (1996). Relevance of HCI guidelines to learning. In *Machine Mediated Learning 5* (2). 119-133. Hillsdale, NJ: Erlbaum.
- Jonassen, D. (1991). Objectivism vs. constructivism. *Educational Technology Research and Development*, 39(3), 5-14.
- Hausmann, R. & VanLehn, K. (2007). Explaining Self-Explaining: A Contrast Between Content and Generation. *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, pp. 417—424.
- 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.
- Koedinger, K. R. & Alevan, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*, 19 (3): 239-264.
- Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2008). Why tutored problem solving may be better than example study: Theoretical implications from a Simulated-Student study. To appear in *Proceedings of the 9th International Conference on Intelligent Tutoring System*, Montreal, June, 2008.
- Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14-19.
- McLaren, B.M., Lim, S., & Koedinger, K.R. (2008). When is assistance helpful to learning? Results in combining worked examples and intelligent tutoring. To appear in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, Montreal, June, 2008.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on papers in this symposium. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 283-310). New York: Academic Press.
- Paas, F.G.W.C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Ed. Psych.*, 84, 429-434.
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K. R., McDaniel, M., et al. (2007). *Organizing Instruction and Study to Improve Student Learning*. (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 3-8.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1051-1057.
- Pavlik Jr., P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29(4), 559-586.
- Pavlik Jr., P. I., & Anderson, J. R. (accepted). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*.
- Pavlik Jr., P. I., Bolster, T., Wu, S., Koedinger, K. R., & MacWhinney, B. (2008). Using optimally selected drill practice to train basic facts. To appear in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, Montreal, June, 2008.
- Ringenberg, M., VanLehn, K. (2006). Scaffolding problem solving with annotated, worked out examples to promote deep learning. *8th International Conference on Intelligent Tutoring System*.
- Schmidt, R.A. & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207-217.
- Schwonke, R., Wittwer, J., Alevan, V., Salden, R.J.C.M., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? *Proceedings of the 2nd European Cognitive Science Conference* (pp. 59-64).
- Sweller, J., Van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.
- Stark, R., Gruber, H., Renkl, A. & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? (Does the combination of worked-out examples and problem-solving tasks pay off?) *Zeitschrift für Pädagogische Psychologie*, 14(4), 206-218.
- Sweller, J. & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction* 2, 59-89.
- Trafton, J.G. & Reiser, B.J. (1993). The contributions of studying examples and solving problems to skill acquisition. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1017-1022).
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.