The Role of Struggle and Productive Failure In Learner Assistance

Dar-Wei Chen

School of Psychology
Georgia Institute of Technology

The Role of Struggle and Productive Failure In Learner Assistance

Preliminary exam presented to:

Dr. Richard Catrambone, advisor
School of Psychology
Georgia Institute of Technology

Dr. Phillip Ackerman
School of Psychology
Georgia Institute of Technology

Dr. Jamie Gorman
School of Psychology
Georgia Institute of Technology

Dr. Mark Guzdial
College of Computing
Georgia Institute of Technology

Dr. Rick Thomas
School of Psychology
Georgia Institute of Technology

Table of Contents

List of Tables

EXECUTIVE SUMMARY

A question in education that has been extensively researched concerns the appropriate amount of guidance for students. According to one school of thought, learners should receive relatively little guidance and instead be allowed to freely explore the problem space without many constraints, enabling them to connect intuitively and actively with the environment; such methods are often called “minimal guidance” (MG) methods. In contrast to MG methods are what might be characterized as traditional or “direct instruction” (DI) methods, which limit the amount of exploration and instead provide concepts and canonical solutions early in the learning process. The DI school of thought argues that providing complete information to learners early on is the best way to ensure accurate mental schemas, low mental workload, and less frustration.

Existing literature tends to portray this debate as a choice between two mutually-exclusive options in which one must be better. However, both methods are, in fact, important in education (Schwartz & Martin, 2004). Furthermore, the debate is not so simple for a variety of reasons. For example, in the realm of individual differences, it is likely that learners of higher pre-existing ability will perform more strongly in environments with low structure because they are more apt to relate new information with prior knowledge (Peterson, 1987), a hallmark MG-related behavior. Low-ability learners are more likely to need help developing strategies and tend to learn using rote memorization, so high structure tends to be favorable (Snow, 1982). Regarding the difficulty of materials: The discrepancy reduction model (Dunlosky & Hertzog, 1998) posits that learners are most motivated to study items that they are furthest away from mastering, while the region of proximal learning framework (Metcalfe, 2002) hypothesizes instead that learners tend to attempt items just out of their reach for a higher learning return. Lastly, many MG and DI methods are similar in using scaffolding to reduce the amount of struggle a learner must endure, instead of leveraging struggle for greater learning.

An example of a MG-based method that leverages struggle (and achieves both the benefits usually associated with MG along with those of DI) is “productive failure” (PF), a relatively new method positing that learners benefit from failing early on in learning (e.g. Kapur, 2008; Schank, 1997). In PF, students start a learning session by solving ill-structured problems (“generation period”) before receiving canonical solutions (“consolidation period”). During the generation period, scaffolding is usually withheld, so learners in PF environments often struggle and fail when tested immediately after generation. However, in assessments after consolidation, PF learners demonstrate a “latent productivity” in initial failure (Kapur, 2008, p. 379).

Productive failure is hypothesized to help learners in a variety of ways that DI methods do not. According to Chi’s (2000) “imperfect mental model” theory, learners must be able to recognize flaws in their initial mental models before being able to repair them; after the failure that almost certainly happens during the PF generation periods, learners are more likely to identify and reflect on these flaws to understand why they occurred (leading to better context transferability), as well as study related feedback that they receive (Kulhavy & Stock, 1989). DI methods tend to produce shallower learning because problem-solving is merely a re-activation of prior information. In terms of cognitive load, learners in PF conditions mitigate working memory constraints during the generation period by relying heavily on the activation of prior knowledge (Kapur & Bielaczyc, 2011), which also provides frameworks for deeper schema encoding.

Eighteen papers that included a direct PF manipulation were examined in a meta-analysis. In general, it appears that PF methods are superior to DI methods on conceptual tasks (performance for PF learners is 0.66 SD higher, on average, than those using DI) and far-transfer tasks (0.49 SD) for which MG is typically more suited, while the two methods are relatively equal on procedural and near-transfer tasks (0.06 SD) for which DI is typically more suited. The analysis also suggests that PF is more effective without generation period scaffolding and when undertaken individually, and that “vicarious failure” is not as effective as actual failure.

In future investigations of PF, two directions are important and intriguing, the first of which is scaffolding-PF interactions. Although the meta-analysis revealed PF to be more effective without scaffolding, guidance is likely beneficial for discovery-based methods (e.g., Anthony, 1973), and the relatively small number of studies did not cover many of the possible scaffolding methods to be used. Subgoals and training wheels are two scaffolding methods that allow room for failure but might provide just enough guidance to ensure the failures are productive. The second direction is domain-PF interactions: It is hypothesized that domains in which rote memorization is key (e.g., introductory biology) might not be suitable for PF, but that creative and transfer-heavy domains (e.g., music composition) are more appropriate.

Issues with PF methods also require investigation. For example, the “IKEA effect” states that learners could prefer (or at least be influenced by) their own generated solutions, even if the solutions are shown to be suboptimal (Mochon et al., 2012; Johnson & Seifert, 1994). The hard-easy effect (learners studying difficult materials tend to be overconfident in their understanding; Lichtenstein et al., 1982) could also hinder PF because material presented in MG methods tends to be rated as more difficult. This perception of difficulty is likely driven by perceived cognitive load (Reynolds & Caperton, 2011), another area that requires more study given that both MG and DI schools of thought hypothesize relatively low loads. Even the necessity of failures can be questioned; if the benefits associated with failure can be induced in learners without actual failure, such methods might be preferred by learners (Clark, 1982). Although PF research is relatively young, current and future findings can be used to improve instruction of all types.

CHAPTER ONE

The assistance dilemma

Of the many issues that face researchers today in instructional science, perhaps one of the most essential is what might be termed the “assistance dilemma,” which asks: “How should learning environments balance information or assistance giving and withholding to achieve optimal student learning?” (Koedinger & Aleven, 2007, p. 239). Researchers have debated this seemingly simple question for many years and those that posit answers have sometimes attracted significant followings, but comprehensive answers in the literature have proven elusive.

In general, most answers to the assistance dilemma can be categorized into two schools of thought regarding instructional design: “Minimal guidance” (MG) and “direct instruction” (DI). Environments with minimal guidance are ones in which learners are not directly presented with canonical solutions and full information, but are instead asked to discover this information and create their own solutions to problems. In contrast, direct instruction methods aim to provide students with all of the necessary information and explanations so that they can learn as quickly and efficiently as possible.

Experiments in this area of literature often compare these two types of methods against each other in an effort to simply reveal the superior method. However, in reality, both methods are important, and the comparisons are complex because individual differences play large roles in the effectiveness of each method, and even prominent learning models can diverge in their hypothesized results of the methods. Furthermore, MG and DI methods are often similar in one fundamental aspect: Neither incorporates struggle and failure into the learning process to the extent that it should be incorporated, instead opting for various scaffolding methods to reduce the amount of struggling a learner must endure.

The types of benefits afforded to learners in MG and DI are both important, as are the benefits associated with struggle. Recently, a growing amount of evidence indicates that the “productive failure” (PF) method of instruction (e.g., Kapur, 2008) is an example of a method that could enable students to acquire all of the aforementioned types of benefits while explicitly leveraging learner struggles and failures. This paper will explore the learner assistance debate as it stands now and what productive failure reveals, in a cognitive sense, about the best ways to help students learn.

1.1 “Minimal guidance” school of thought

Minimal guidance methods are all generally based on similar principles, but vary in terms of theoretical reasoning, objectives, and/or implementation.

1.1.1 Discovery learning. One of the earliest iterations of minimal guidance is discovery learning. Broadly speaking, discovery learning is characterized by requiring learners to unearth the rules governing a domain or material for themselves so that they can create new insights (Anthony, 1973). One of the goals of this type of method is to turn students into “as autonomous and self-propelled a thinker as we can – one who will go along on his own after formal schooling has ended” (Bruner, 1961, p. 2). Therefore, it appears that even early on, the importance of developing intrinsic motivation in learners was evident, as opposed to conditioning learners to respond just to extrinsic motivational factors such as grades. Bruner (1961) also writes that teaching in a way conducive to long-term learning is effective because “when behavior is long-range and competence-oriented, it comes under the control of more complex cognitive structures,” freeing the behavior from “immediate stimulus control” (p. 6). Apparently, minimal guidance methods were being designed with an eye toward retention (not merely aiming to improve performance in the short term), and studies like one carried out by Dean and Kuhn

(2008) have indeed demonstrated that direct instruction is not very conducive to retention of knowledge over time and transfer to novel situations.

1.1.2 Constructivism. Two similar but slightly different approaches use the constructivism label: the cognitive theory and the educational philosophy (Guzdial, 1997). Constructivism as a cognitive theory emphasizes the notion that learners should actively construct meanings for concepts; however, a slight conceptual difference from discovery learning is that instead of constructing their own representations (a la discovery learning), learners in constructivism (cognitive) are often merely building “conceptually functional representations of the external world” that are not necessarily unique (Jonassen, 1991, p. 61). That is, constructivism as a cognitive theory holds that common knowledge exists, and that learning environments should be created in a way that allows learners to explore that common knowledge. As an educational philosophy, constructivism is more like discovery learning in that the meanings attached to concepts are considered to be unique to the individual learner (Guzdial, 1997). This emphasis on individualized meaning leads to the idea that learning should happen in contexts that are as close to the real world as possible so that the connections drawn by the learners are more inherently meaningful, not imposed by sometimes arbitrary and isolated school environments (Jonassen, 1991). Supporting constructivism as an educational philosophy are models such as situated cognition (e.g. Brown, Collins, & Duguid, 1988), which suggest that learned information is always associated with the context it was learned in; therefore, providing relevant contexts is key to effective learning. However, one potential issue arises if concepts are thought only to have meaning as constructed by a learner: Does absolute truth exist, and is it sensible to grade learners based on standardized objectives (Guzdial, 1997)?

1.1.3 Impasse-driven learning. Discovery learning and constructivism are methods that do not directly implement learner struggle and failure, per se, instead merely guiding students into more active roles while learning. One MG method that does implement struggle/failure to some extent is impasse-driven learning; impasses are defined by VanLehn at al. (2003, p. 220) as situations in which a student is stuck, “detects an error, or does an action correctly but expresses uncertainty about it.” Impasse-driven learning is governed by the principle that students learn more effectively when reaching impasses because cognitively, they are more likely to adopt learning orientations than if safeguarded against impasses (VanLehn et al., 2003). These learning orientations drive the students to discover what they do not understand, whether that involves searching memory, examining the environment, or asking nearby people (VanLehn et al., 2003). Conceptually, this method is similar to breakdown-driven learning (Winograd & Flores, 1987), which depicts learners reacting to misunderstandings by moving a concept from “ready-to-hand” (assuming a concept is understood and using it to proceed) to “present-at-hand” (consciously reflecting on it to resolve a misunderstanding).

However, in impasse-driven learning, tutors are asked to step in to provide explanations once a student fails to provide his or her own explanation, which could be too soon for learner struggle and failure to achieve its full benefits. Cope and Simmons (1994) found that immediate feedback can sometimes turn into a crutch that shields learners from having to create high-level problem-solving strategies, while those that are forced to wander a bit without feedback do exhibit such strategies. Therefore, although impasse-driven learning is a step in the right direction regarding its utilization of struggle and failure, it might not be enough.

1.1.4 Other general arguments for minimal guidance. MG methods generally are also hypothesized to afford these benefits to learners:

- By requiring learners to rely chiefly on connections to prior long-term knowledge during problem-solving, MG methods can help mitigate working memory constraints (Kapur & Bielaczyc, 2011). Learners in DI environments tend not to rely on long-term knowledge as much because the new information is more readily-available for access than in MG environments. These connections to long-term knowledge in MG environments also serve to ensure that the new material is understood at a deeper level than information that might just be held in working memory. Understanding material at this deeper level increases storage strength, a better indicator of enduring learning than retrieval strength, which reflects current activation and accessibility and is often dependent on cues and context (Bjork & Bjork, 1992). Therefore, although MG methods might produce slower initial learning as students struggle without much scaffolding, the learning is more likely to be permanent and stable.
- Students using MG methods are more likely to learn how to structure complex problems independently, which is a trait indicative of expertise (Chi, Feltovich, & Glaser, 1981). When using DI methods, students can lean on the initial presentation of canonical instructions to help them during problem-solving, leading to the possibility that they never learn how to structure the problems by themselves. To the extent that this hypothesis is true, MG methods are expected to produce more expertise than DI methods.

1.2 "Direct instruction" school of thought

Opposite minimal guidance in the debate over learner assistance are direct instruction methods, which generally posit that students will learn most effectively when strongly guided, as opposed to finding their own way in MG environments.

1.2.1 The "worked-example" effect and working memory. According to Kirschner, Sweller, and Clark (2006), the worked example is "the epitome of strongly guided instruction" and "provides some of the strongest evidence for the superiority of directly guided instruction over minimal guidance" (p. 80). The effect has been shown to occur, for example, when students learn algebra, as the students who studied worked examples performed better than those who solved problems (Sweller & Cooper, 1985). Driving the worked-example effect is the organization of the material such that problem-solving search is streamlined and attention is properly directed, leading to lower working memory load than if the learner used a minimal-guidance method (Kirschner, Sweller, & Clark, 2006). This streamlining of the search process is important because the search strategies that most learners use, particularly novices (who lack schemas to integrate new information with prior knowledge), are often inappropriate for constructing durable schema (Rourke & Sweller, 2009). More specifically, novices often resort to processes such as trial-and-error or means-ends analysis, which heavily burden working memory such that it is not available for contributions to long-term memory (Kirschner, Sweller, & Clark, 2006). Because all conscious processing happens in working memory, students could potentially search problem spaces for extended periods of time without altering long-term memory (Sweller, Mawer, & Howe, 1982), and moreover, the purpose of problem-solving search is to find solutions, and it is therefore inefficient in altering long-term memory (Sweller, 1988). In a practical sense, most people in educational settings are learning material that they do not have much experience with, so instructional methods that help novices are the ones most likely to be useful in school environments.

Interestingly, both schools of thought (minimal guidance and direct instruction) hypothesize lower working memory loads for learners using their methods, albeit for different reasons. It is posited that MG methods induce lower working memory load because of learners' reliance on long-term knowledge (e.g., Kapur & Bielaczyc, 2011), while data supporting methods such as worked examples suggest that reductions in problem-solving search are more effective in reducing working memory load (Kirschner, Sweller, & Clark, 2006).

1.2.2 Other examples and arguments in favor of direct instruction. Worked examples are just one instance of direct instruction; lectures, modeling, videos, presentations, demonstrations, and discussions can all serve as DI methods as well (Clark, Kirschner, & Sweller, 2012). All DI methods are theorized to reduce the misconceptions that learners often develop when receiving minimal guidance (Brown & Campione, 1994), and be more efficient as well due to the fact that mentally-taxing "false starts" in understanding are less inherent than in MG (Carlson, Lundy, & Schneider, 1992). That is, DI reduces the probability of encoding errors and facilitates the development of correct domain knowledge better than MG does (Sweller & Chandler, 1991).

As opposed to MG methods in which problem-solving is used as a vehicle for delivering content, DI methods typically use problem-solving as a way to allow learners the opportunity to practice applying the canonical instruction they learned initially, and scaffolding during problem-solving is often "faded" (i.e., gradually withdrawn) over time in the hopes that learners can eventually solve the problems on their own. This arrangement is hypothesized to be more optimal than MG arrangements because early on in problem-based MG learning, students can sometimes have trouble distinguishing learned content knowledge from the specific problems in which they learned the content (Patel, Groen, & Norman, 1993); therefore, problem-based MG approaches could increase the possibility of learners incorrectly applying knowledge from old contexts to problems in new contexts. In contrast, DI methods tend not to expose learners to problems until the learners have more experience with the domain and can properly contextualize the information (Kirschner, Sweller, & Clark, 2006).

According to Mayer (2004), direct instruction has proven to be pedagogically superior many times throughout the years, regardless of the minimal guidance method serving as the comparison group. It is also hypothesized to be superior on secondary measures such as disengagement and frustration in learners; that is, “problem-solving prior” methods are likely to induce greater learner disengagement and frustration than DI methods (Hardiman, Pollatsek, & Weil, 1986).

1.3 Role of scaffolding in the assistance dilemma

Although minimal guidance and direct instruction are at odds regarding many issues, as demonstrated previously, one issue that is agreed upon to a large extent is the role of scaffolding. That is, most MG and DI methods implement some amount of scaffolding, presumably due to the belief that learners will fail to accomplish the necessary learning if not scaffolded properly (e.g., Aulls, 2002; Kirschner, Sweller, & Clark, 2006). One of the objectives of scaffolding in many MG and DI methods is to protect learners from struggle and failure.

Scaffolding methods can be defined as processes that enable learners to “solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts” (Wood, Bruner, & Ross, 1976, p. 90). Reiser (2004) observes that scaffolding helps learners mainly by enabling them to key on essential goals and concepts so they do not proceed mindlessly and/or fruitlessly through the task; Pea (2004) calls that property "channeling and focusing" while adding that modeling solutions for a task can also be one of the properties and functions of scaffolding.

1.3.1 Examples of scaffolding methods. The question of how to scaffold learners is arguably one of the most fundamental questions in the education literature. Due to its importance and the vast possibilities offered to researchers who wish to tackle the question, many have offered solutions. Some are as simple as automating a student's "busy work" such as displaying data, performing calculations, or storing information, which are aimed at reducing extraneous cognitive load (Hmelo-Silver, Duncan, & Chinn, 2007), but others are more complex. Here are some examples of scaffolding methods from existing literature:

- *Training wheels* (e.g., Carroll & Carrithers, 1984; Catrambone & Carroll, 1987): The researchers developed an interface for teaching word processing that blocked learners from troublesome "side tracks and error tangles" (p. 800) during the course of learning, finding that participants using this "training wheels" interface performed better and learned more efficiently than those who had access to the complete word processing system. Instead of direct instruction that tends to relegate learners to a passive role, this method allows learners to actively explore the environment while limiting the wasted time that can sometimes hinder unguided instruction.
- *Subgoals* (e.g., Catrambone, 1998): Decomposing the overall goal of a problem into a series of smaller goals assists learners in forming subgoals that can be helpful in problem-solving performance. Subgoal labels can induce self-explanation in learners (for example, why particular steps should be grouped together into a subgoal) that steer them away from simply memorizing superficial

details and steps but understanding the functional reasons for a grouping of steps. The deep structure provided by subgoals enables learners to transfer their learning to new problems that involve similar subgoals (but perhaps new steps within those subgoals).

- *Technology-driven reflection* (Lin, Hmelo, Kinzer, & Secules, 1999): A framework developed by Lin and colleagues outlines four ways that technology is particularly suitable for in terms of providing scaffolding for reflection.
    - *Process displays* enable “normally tacit learning processes” to be seen in “explicit and overt” ways (p. 47). Showing the students’ work to them allows for reflection to be much more accessible because hidden/covert processes are difficult to analyze.
    - *Process prompting* facilitates students’ explaining and evaluating “what they do before, during or after problem-solving acts” (p. 49), which helps them monitor their learning progress.
    - *Process modeling* encourages learners to use experts’ solutions as models for their own solutions via tracking, replaying, and analyzing of expert thinking processes. This method is similar to the “cognitive apprenticeship” learning method, which emphasizes modeling, performance reflection, coaching, and overt articulation of tacit processes as also described in this Lin et al. framework (Collins, 1991).
    - *Social discourse* can be used as a mechanism for A) students to learn from various perspectives, and B) social factors to be leveraged as a motivating force to produce quality work.

- *Evidence mapping* (Toth, Suthers, & Lesgold, 2002): When learning to evaluate empirical evidence in the face of competing hypotheses, graphically representing the relationships between pieces of evidence produces superior reasoning skills than representing the evidence through familiar methods such as prose writing. The perceptual salience of relationships is much higher when evidence is mapped graphically; writing prose does not inherently require learners to focus on important information and therefore can lead to wasted effort on “extra information” (p. 281).
- *Collaboration*: One of the most powerful ways for students to learn is to interact with their peers and receive different perspectives, something that canonical instruction cannot always provide. Roschelle (1992) observes that conversational interaction can produce “convergent conceptual change,” in which learners’ individual interpretations of concepts “converged towards shared knowledge,” and discrepancies are easily recognized and repaired through discussion. However, facilitating collaboration is often necessary to ensure effective discourse. An example of a facilitative mechanism is the collaboration script, which can dictate the general sequence or allocation of learning activities to facilitate discussion and argumentative activities while also limiting off-topic activity (e.g., Stegmann, Weinberger, & Fischer, 2007; Kobbe et al., 2007). More specific ways to facilitate collaboration include prompts and sentence starters (Nusbaum et al., 2002) as well as input text fields (Kollar et al., 2005), which are all effective in requiring logical flows for discussion. Collaboration facilitation is likely most effective for deeper elaboration, which does not often happen

spontaneously in groups (King, 2007). Dillenbourg (2002) cautions, however, against over-scripting collaborative activities, which could reduce motivation and suppress connections to learners' previous knowledge.

- *Knowledge integration framework* (Linn, 2000): Interestingly, the aim of this framework is to guide learners heavily, but achieve goals traditionally associated with minimal-guidance methods (e.g., integrating information with pre-existing knowledge). The four elements of this framework are:
    - *Identifying new goals for learning*: Instructors should A) select models of novel concepts that build on the intuitive ideas of their students, B) emphasize the integration of ideas instead of isolating them, and C) provide resources for students to use after the course
    - *Making thinking visible*: Dynamic models that illustrate thought processes can be very helpful to learners in recovering from unsuccessful forays
    - *Encouraging autonomous learning*: Students should learn to evaluate evidence on their own instead of relying on the instructor to dictate the correct answer, which could involve "renegotiating the authority structure in most classrooms" (p. 6)
    - *Providing social supports*: Common impediments to be overcome in group work include A) the exclusion of others by a sub-group, B) interruptions and insults of other group members, C) "natural leaders" gaining more from group experience than others, and D) the division of tasks into parts and having each person work alone on his or her part

1.3.2 Classifying methods by scaffolding implementation. The amount and types of scaffolding in a given method make up some of the differentiating characteristics between minimal guidance and direct instruction, but classifying a method given its scaffolding levels has not always been consistent in existing literature. These arguments are not just semantic or academic – any discussion about the merits of MG and DI methods can only proceed properly with suitable definitions of each school of thought.

Inquiry learning (IL) is an example of a method that is commonly thought of as one with minimal guidance, characterized by Rutherford (1964) as implementing the teaching of inquiry as content (i.e. learning how to carry out scientific inquiry) or inquiry as technique (i.e. learning scientific principles using inquiry). A more modern but related method that has been debated is problem-based learning (PBL), a method commonly used in medical schools. In applications of PBL to medical education, students diagnose and suggest treatments for given symptoms with supervision and intermittent instruction from faculty (Barrows & Tamblyn, 1980). Kirschner et al. (2006) classify both problem-based learning and inquiry learning as MG methods because fundamentally, students are asked to discover solutions for themselves in authentic environments ("no clear-cut distinguishing features" between PBL and IL; Hmelo-Silver et al., 2007, p. 100). However, Hmelo-Silver and colleagues (2007) disagree, stating that the level of scaffolding provided in PBL and IL is enough to disqualify them from being labeled as minimally-guided methods; the mini-lectures and benchmark lessons often presented to students on a "just-in-time" basis are direct instruction activities, just with timing that is not traditionally associated with DI methods. PBL and IL implement just-in-time direct instruction to help students know the reasons why information is needed at a given point, as opposed to merely memorizing information in a rote manner (Edelson, 2001). With confusion about how methods should even be classified, it is not surprising that solutions for learner assistance questions are not very definitive.

CHAPTER TWO

External factors influencing the effectiveness of methods

Existing literature in the learner assistance realm often portrays minimal guidance and direct instruction as methods in direct opposition to each other on many levels. However, not only should instructors and researchers refrain from choosing one method over the other, given that both have a place in education (e.g., Schwartz & Martin, 2004), individual differences and competing learning models can make the decision more nuanced than the literature sometimes suggests. Furthermore, productive failure methods (described in chapter three) have shown some promise in helping learners obtain both the advantages usually associated with MG and those with DI, implying that the currently-constituted learner assistance debate might be a false choice.

2.1 Individual differences

Any given learning environment is likely to have students of varying capabilities, so accounting for individual differences is essential to creating an environment that maximizes learning potential.

2.1.1 General intelligence. Broadly speaking, MG learning environments are less structured than DI learning environments (e.g., in response to a learner error, a highly-structured method might provide explicit directions for correcting an error, while a less-structured method might merely notify the learner of the error's existence; McLaughlin, Rogers, & Fisk, 2006), and according to Snow (1989), high-intelligence learners tend to perform relatively well in low-structure environments because of the "metacognitive and self-regulatory skills associated" with high intelligence. Sullivan and Skanes (1971) add that learners of high intelligence transfer more effectively, a skill that lends itself well to low-structure environments where learners must "fill in gaps" themselves, which is itself a transfer skill (Snow, 1989, p. 22). Furthermore, high-intelligence learners often suffer decreased performance with high amounts of structure because they are not able to use their own developed methods as much (Gray, 1982). Conversely, low-intelligence learners usually perform well with higher degrees of structure in their learning procedures because they are more likely to need help developing strategies, and rote memorization is a larger part of their learning habits (Snow, 1982). Supporting all of these conclusions is the finding that high-intelligence learners achieve more with less-guided heuristic-based instructions, while low-intelligence learners perform better with heavily-guided algorithmic instruction (De Leeuw, 1983).

The structure level of a method can sometimes be seen in the way it uses problem-solving; minimally-guided methods often implement problem-solving as relatively large parts of the learning process (especially early on in learning), and some research shows that this aspect of MG methods is also favorable to learners of high intelligence. Hsu, Kalyuga, and Sweller (2015) have found that problem-solving induces relatively low cognitive load in learners of high intelligence, so heavy guidance before problems is less necessary for them to achieve learning objectives. Those with high intelligence also tend to relate new information to pre-existing knowledge more effectively (Peterson, 1987), an ability that is magnified in importance during MG learning because pre-existing knowledge is essential for creating solutions when guidance is sparse. Conversely, low-intelligence learners generally do require guidance before tackling unstructured items (such as problems) because the cognitive load associated with problem-solving would otherwise be very difficult to overcome (Koedinger, Corbett, & Perfetti, 2012).

In short, teachers with students of high pre-existing abilities can generally afford to allow those students to explore in relatively free form, and they can also feel confident in the effectiveness of using problem-solving as a mechanism for learning. Students with lower pre-existing abilities will learn more successfully in well-structured environments and when problem-solving is implemented after canonical instruction (as opposed to before).

2.1.2 Pre-existing knowledge (or domain experience). The extent of a learner's domain experience also impacts the success of a given instructional method. For example, during ill-structured problem-solving, novice learners often cannot identify relevant knowledge gaps because they do not have the requisite mental structures and schemas in place as reference, while learners experienced in the domain are more able to fit the new material into proper contexts (van Gog, Kester, & Paas, 2011). Therefore, experienced learners are more likely to benefit from the use of problem-solving as a learning mechanism (as is often done in MG) because they "know what they don't know" and can deliberately work to address those weaknesses. A finding that corroborates these notions is the fact that novice learners, when tasked with solving problems, are relatively less able to assess their performances on those problems when compared with experienced learners (Dunning et al., 2003). This connection is intuitive – learners that are not as able to identify gaps in their knowledge should also be expected to lack the wherewithal to accurately gauge their comprehension during problem-solving periods. As a result, any method relying on problem-solving as a key content delivery mechanism (i.e., many MG methods) is not well-suited for novice learners. For these reasons, younger kids, who tend to have little prior experience to draw from in any given domain, might not reap the benefits of MG methods, although facilitative prompts could help (Mazziotti, Loibl, & Rummel, 2015).

Domain experience also interacts with text coherence, "the extent to which the relationships between ideas in a text are explicit" – examples of increasing and decreasing text coherence include using or not using ambiguous pronouns, relatable descriptions, transitions and connectives between ideas, topic headers, and topic sentences (McNamara, 2001). Data indicate that learners with less domain knowledge benefit from reading high-coherence texts, while high-knowledge learners are better served reading low-coherence texts, particularly on conceptual measures such as inference-making (McNamara & Kintsch, 1996); this phenomenon likely occurs because low-coherence texts force readers to process presented ideas in relation to prior knowledge and in more active ways, both of which domain-experienced readers are more likely to be able to do with their pre-existing knowledge (McNamara, 2001). These findings regarding text coherence seem to mirror the previously-discussed findings regarding domain experience and learner assistance; that is, to the extent that low-coherence texts resemble minimal guidance and high-coherence texts resemble direct instruction, it is only logical to posit that experienced learners will perform well with MG and novice learners with DI. The idea that domain-experienced learners suffer with increased guidance has also been termed the "expertise reversal effect" (e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003), which hypothesizes that cross-referencing existing mental schemas with the newly-provided instruction can create strain on working memory and cognitive overload, even if the learner recognizes the redundancy of the information. Some researchers might warn against the dangers of implementing MG methods because of the tendency for minimal guidance to produce slower initial learning – which, in turn, could discourage learners – but that fear might be overstated if MG is primarily applied to experienced learners: Students with domain experience are usually more engaged in learning and interested in creating learning strategies than novice learners, so any discouraging initial results are less likely to be impactful (Alexander, 2003).

Expertise reversal can also affect the impacts of various types of scaffolding methods as well, which in turn can differentially affect lightly-scaffolded MG methods and highly-scaffolded DI methods. For example, the imagination effect postulates that students who imagine the content and procedures of instruction before learning the material perform better than those who simply learn the material (Cooper, Tindall-Ford, Chandler, & Sweller, 2001); this effect is much more pronounced for learners with pre-existing knowledge of the domain because their schemas are more cohesive (i.e., less processing of individual elements) and complete, while inexperienced learners must cope with processing elements individually (high cognitive load) and possibly incorrect schemas (Kalyuga et al., 2003).

On a more general level, high intelligence and high pre-existing knowledge are similar in their interactive effects on the effectiveness of learner assistance levels: the higher a student's general intelligence or pre-existing knowledge, the less guidance that student will require (and the lower the intelligence or knowledge levels, the more guidance is necessary). However, the mechanisms driving these phenomena are somewhat different. Regarding intelligence, self-regulation, creativity, and cognitive load issues seem to be most important in determining optimal assistance levels; for domain experience, schemas are central, along with self-regulation as well.

2.1.3 Other learner characteristics. In addition to general intelligence and domain experience, other learner characteristics can affect the use of structure in learning environments.

- *Assessment expectations*: If a student is expecting a free-recall test after the learning period, he or she will tend to examine the structure of the text more closely than if expecting a multiple-choice test (d'Ydewalle, 1984). Presumably, the effects of MG and DI will be more pronounced when learners are expecting to reproduce information (free recall) than when expecting to perform recognition tasks (multiple-choice) because the impact of text structure will be larger.

- *Anxiety*: High-structure environments tend to help anxious and/or conforming students, while independent and non-anxious students are relatively better off in low-structure environments (Snow, 1989). Interestingly, high-ability learners tend to be most negatively affected by anxiety, while low-ability learners are often aided by slight anxiety (Hagtvet, 1986); therefore, even the structure-intelligence interactions described previously in this document might not be as simple as previously described.
- *Learner intentions*: Sometimes, the degree of structure in a task can be surpassed in importance by learner intentions. For example, if a learner intends to make personal connections to materials, he or she will process information more deeply; learners that intend to merely regurgitate for an upcoming test will tend to learn at a surface level (Entwistle, 1985). Alexander (2003) notes that "…even though knowledge and strategies remain keys…individuals' investment in their learning and development is equally critical" (p. 12).
- *Autonomy*: Self-determination theory (Deci & Ryan, 2008) posits two types of motivation: Autonomous ("behaving with a full sense of volition and choice") and controlled ("behaving with the experience of pressure and demand toward specific outcomes that comes from forces perceived to be external to the self"). Depending on a learner's disposition with regards to these two types of motivations, he or she might be more suitable for minimal guidance or direct instruction; in particular, those with a "self-determined motivational orientation" could prefer MG environments in which autonomy is afforded to the learner in relatively large quantities, while those that tend to be extrinsically-motivated could require the

structure and external pressures of DI environments to perform well (Reynolds & Caperton, 2011).

2.2 Hypotheses from two prominent learning models

Competing learning models hypothesize different outcomes for various levels of learner assistance and learners' pre-existing abilities/knowledge.

2.2.1 Region of proximal learning. This theory posits that learners are more likely to allocate mental resources to material that they perceive to be just outside their grasp because of the relatively high learning return probability (compared to very difficult materials), while difficult materials are only attempted if extra time is available (Metcalfe, 2002). A similar theory is the zone of tolerable problematicity, which suggests that every learner has a perceived difficulty range within which he or she is most likely to engage problems in productive ways: Below this range, the learner can do things relatively automatically without thinking and errors are due mostly to cognitive slips (not conceptual misunderstandings); above this range, learners must improvise too much and errors are more likely to arise from conceptual misunderstanding, causing decreases in motivation and increases in anxiety (Elshout, 1985).

To the extent that the structure level of materials is inversely proportional to its perceived difficulty, it is possible to hypothesize that learners of relatively high ability and knowledge will find the difficulty levels of MG environments to be relatively nearer to the top end of their ranges (at least when compared to DI environments), and according to the aforementioned learning models, they will be more motivated to learn in these situations. Learners of relatively low ability and pre-existing knowledge might find that in MG and without the scaffolding of direct instruction, the material's difficulty seems to be too far above their ranges and therefore yields discouragingly low probabilities of learning return. Conversely, in highly-structured environments, low-ability learners will likely enjoy the scaffolding that brings the difficulty level closer to their ranges, but the same activity would become dull for high-ability learners. In a sense, these learning models suggest the existence of a “sweet spot” for the difficulty of materials, but this sweet spot appears to differ from one learner to the next. This notion is supported by various curvilinear models of arousal indicating that performance suffers with too much or too little arousal (e.g., Lens, 1983). However, whether learners can accurately perceive the difficulty of material and judge their learning is a debated topic. Some research suggests that some learners cannot be trusted to do it (e.g., “pseudo-experts,” those who self-report expertise much higher than their actual expertise, often allocate study time as if they are actual experts; Metcalfe, 2002), but other research indicates that they can be trusted (e.g., participants scored more highly when allowed to study the materials they requested to study again, as opposed to studying the materials they did not request; Kornell & Metcalfe, 2006).

2.2.2 Discrepancy reduction model. In contrast to the previously-described theories, the discrepancy reduction model postulates that learners are most motivated to study the items that they are furthest away from mastering (i.e., where they have the largest discrepancies between current state and goal state; Dunlosky & Hertzog, 1998). Supporting this model are findings such as ones in which learners who are told to activate particular items in working memory spend less time studying items related to those activated items (e.g., Machiels-Bongaerts et al., 1993)

As for how these concepts relate to the assistance dilemma and individual differences, it appears that the discrepancy reduction model would favor minimal guidance for all learners, regardless of abilities and knowledge levels, because it would create the largest perceived discrepancy between current state and goal state. That is, learners are faced with more struggles during low-structure tasks such as open-ended problem-solving (especially early on in the learning process), and these challenges to achieve the goal state are more motivating than the comparatively tame challenges of direct instruction. This hypothesis stands in contrast to the hypothesis of theories such as the previously-discussed region of proximal learning, which states that minimal guidance would be most effective only for learners of high ability and pre-existing knowledge (direct instruction being most effective for learners of low ability and pre-existing knowledge).

One criticism of the discrepancy reduction model is its theoretical recommendation of always providing the most difficult (e.g., unguided) materials to all learners for the purposes of motivation; Metcalfe and Kornell (2005) note that “if people held to this model, and an item were unlearnable, they could, in principle, study for an unlimited amount of time” (“labor in vain”; p. 464) and be perpetually motivated to do it because of the unendingly large gap in understanding. Hypothetically, even low-ability or low-knowledge learners would not require guidance, lest it close the gap and decrease motivation. Because learners are not generally motivated in this way (learners are usually motivated by methods that improve short-term performance, as opposed to those that are relatively worse in short-term but better for long-term and retention; Baddeley & Longman, 1978), discrepancy reduction does not seem to be a tenable model for describing learner interactions with various guidance levels. Instead, given that much of the existing literature generally supports the notions that learners of low ability/knowledge require more guidance and learners of high ability/knowledge learn more effectively with less guidance, the proximal learning theory (and related theories) appears to be a better fit with prevailing findings.

CHAPTER THREE

The "productive failure" paradigm

A relatively large portion of learner assistance research portrays the debate as an "either-or" proposition, implying that either minimal guidance or direct instruction is the most effective way to teach, and that a learner's use of one method largely precludes him or her from the benefits of the other method. Important unanswered questions (e.g., does the quality of an invented solution impact learning?) also complicate the debate (Loibl & Rummel, 2014a). However, both schools of thought offer important advantages to learners, and deciding between the two schools of thought might be a false choice anyway.

Although it is a fairly new method, there is some evidence that productive failure can produce effective student performance (e.g., Kapur, 2011), which will be discussed in this chapter. What the evidence reveals about the nature of learner assistance is that struggle and failure is A) important for students on a cognitive level, and B) able to be leveraged as part of a method that achieves both the benefits usually associated with minimal guidance and those of direct instruction (i.e., it is not the case that only one set of benefits or the other is achievable).

To describe how productive failure is different from typical minimal guidance and direct instruction, it will be instructive to briefly and generally summarize MG and DI methods:

- *Minimal guidance* methods use problem-solving as the chief mechanism for learning, often implementing scaffolding during problem-solving periods to ensure that students learn all of the necessary material and do not stray too far off-track.

- *Direct instruction* methods use canonical instruction as the chief mechanism for learning, often with problem-solving afterward serving as a "practice period" for the students to apply what they have learned.

The guiding principle of productive failure is that struggle early on in the learning process is encouraged and that there often exists "a latent productivity in what initially seemed to be failure" (Kapur, 2008, p. 379). This latent productivity might arise from something like the generation effect, "which refers to the long-term benefit of generating an answer, solution, or procedure versus being presented that answer, solution, or procedure" (Bjork & Bjork, 2011). Like many MG methods, PF uses problem-solving (generation) as the chief mechanism for learning early on; however, instead of scaffolding learners during problem-solving (or using well-structured problems), PF methods generally withhold scaffolding and instruct through ill-structured problems. MG learners are often scaffolded so that they can avoid failure; in PF, failure is purposefully designed into the learning process. In practice, relevant features in PF problem-solving are often deliberately made inconspicuous so learners will be unlikely to guess the canonical solutions and instead forced to lean heavily on their own intuitive ideas and prior knowledge for generated solutions (Loibl & Rummel, 2014a).

Some type of initial assessment is usually implemented after problem-solving to ensure that learners fail concretely, and then canonical instruction follows soon after that for learners to rectify mistakes and misunderstandings before taking the post-tests. The productive failure method aims to help learners reap some of the benefits from intuitive exploration and canonical instruction, both of which are important for learners (Schwartz & Martin, 2004), and that MG or DI alone cannot provide. Table 1 below summarizes prototypical MG, DI, and PF methods and their relationships to each other.

Table 1.

Summary of prototypical minimal guidance, direct instruction, and productive failure methods

<table>
<tr><th></th><th>Period 1</th><th>Period 2</th><th>Period 3</th><th>Period 4</th><th>Period 5</th></tr>
<tr><td>Minimal guidance</td><td colspan="3">Problem-solving (with scaffolding)</td><td>Immediate assessment</td><td>Retention assessment</td></tr>
<tr><td>Direct instruction</td><td colspan="2">Canonical instruction</td><td>Problem-solving (w/ scaffolding)</td><td>Immediate assessment</td><td>Retention assessment</td></tr>
<tr><td>Productive failure</td><td>Problem-solving (no scaffolding)</td><td>Initial assessment</td><td>Canonical instruction</td><td>Immediate assessment</td><td>Retention assessment</td></tr>
</table>

One of the recurring mathematics problems from the productive failure literature can serve as an example to concretely illustrate the differences between PF and DI (from Kapur & Lee, 2009; Kapur & Bielaczyc, 2012); the students learned about the concept of average speed (instructional time was maintained evenly across conditions).

- *PF conditions:* In the first period, students solved an ill-structured problem in which the story was about two people who had to reach a destination at the same time with various constraints (different modes/speeds of transportation, waiting time, etc.). The relevant information was presented indirectly through a dialogue that the students read. After this problem, students were assessed on extension problems based on the ill-structured problem. In the consolidation phase (canonical instruction), the students shared their invented solutions, after which the teacher shared the canonical solutions and general concepts while contrasting student-generated solutions with those canonical solutions.
- *DI conditions:* Students received canonical instruction and examples first from the teacher, then worked on well-structured problems (i.e., relevant information concretely and directly presented) that were similar in nature to the examples they had practiced. An example of one of these problems is “Jack walks at an average

speed of 4 km/hr for one hour. He then cycles 6 km at 12 km/hr. Find his average speed for the whole journey" (Kapur & Bielaczyc, 2012, p. 83).

3.1 The case for productive failure

The following sections outline key arguments for the effectiveness of productive failure as an instructional method.

3.1.1 Intuition plus formal knowledge. As stated in earlier parts of this document, minimal-guidance methods are hypothesized to mitigate working memory constraints by requiring learners to lean heavily on prior knowledge and intuition during problem-solving (Kapur & Bielaczyc, 2011). Even if some of these gains are negated by the fact that searching problem spaces inherently also increases working memory burdens (Sweller, 1988), it leads to better encoding and assembling of schemas (Hiebert & Grouws, 2007), which provide relevant frameworks for future information to build from. Furthermore, the students that are able to persist through the higher cognitive demands of PF often find themselves more engaged because of the greater agency afforded to them in initial problem-solving (diSessa, Hammer, Sherin, & Kolpakowski, 1991).

This activation of prior knowledge structures helps learners create understanding that is more enduring and transferable because of the new material's connection with information that the learner already has a context for (Kapur, 2012). PF methods strive to leverage this intuitive prior knowledge, along with formal knowledge from canonical instruction (period 3), to provide learners with benefits that MG and DI methods alone are not as able to provide. One of those benefits delivered by the blending of intuition and formal knowledge is the ability to generate a large diversity of solutions for novel problems (diSessa & Sherin, 2000), which is also beneficial for learners because diverse solutions sets are a hallmark of the way experts solve complex problems (Clement, 1991; Reif & Larkin, 1991). For the "average speed" problem, examples of different invented solutions include algebraic representations of the story, "brute-force" methods (guessing a distance and adjusting), diagrams, and/or conceptual statements about the variables (Kapur & Bielaczyc, 2012). This type of learning can help students develop procedural flexibility, the capacity to extend learned information to problems in other contexts and conditions (Gorman, Cooke, & Amazeen, 2010). Another related benefit of activating prior knowledge and exploring larger problem spaces (as is likely to happen in PF) is the cognitive priming of students to solve transfer problems using that information, even if that information does not directly help participants for any given initial problem (Bransford & Schwartz, 1999).

Of course, many DI methods also implement both canonical instruction and problem-solving, so one might ask whether those DI methods would help learners achieve the same benefits of PF as described above. The key difference lies in the function of problem-solving. In PF, students use problem-solving to "assemble or structure key ideas and concepts while attempting to represent and solve the ill-structured problems" (Kapur, Dickson, & Yhing, 2010, p. 1722). However, DI methods usually implement problem-solving "not as vehicles for making discoveries, but as a means of *practicing* recently-learned content and skills" (Clark, Kirschner, & Sweller, 2012, p. 6). Therefore, when students learn from direct instruction, they are not blending intuition and formal knowledge per se (and receiving the associated benefits of doing so), but receiving formal knowledge and then likely non-creatively reactivating that formal knowledge to solve problems, leading to skills that are not as transferable. Productive failure, in contrast, requires students to first use intuition during problem-solving before receiving canonical instruction as a way to remedy misunderstandings, so both types of knowledge are indeed activated for the purposes of completing post-tests. It is this particular "problem-solving prior" order that often influences learners into adopting expertise- and mastery-oriented goals, which are more conducive to producing durable learning than surface-level learning goals (Belenky & Nokes-Malach, 2012) because of the deep structural knowledge needed for expertise and mastery. Generating solutions before canonical instruction also serves to help students understand the limits of their prior intuitions (DeCaro & Rittle-Johnson, 2012), which then enables them to use formal instruction in ways that are more personalized to their needs.

3.1.2 Failure-related cognition. Schank (1997) has written before that people learn best through "expectation failure," which occurs when people expect a particular outcome from material and that outcome fails to occur. Some of the key principles of expectation failure are listed below:

- If the environment does not allow for failure, learners will fail to develop the creativity necessary for attacking novel problems
- People naturally attempt to explain what happened and modify their thinking accordingly so that they are not continually surprised by similar events
- This expectation failure must occur during problem-solving for it to create reminders that can be activated during future problem-solving

These episodes of expectation failure are important because learners might not otherwise be aware of gaps in their understanding. Learners often exhibit stability bias, the overconfident belief that current accessibility of information (i.e., short-term performance) is relatively stable and will hold in the future (Kornell & Bjork, 2009). Failures, like those induced in PF, help to disrupt this bias because learners are more likely to focus on knowledge gaps after failures, often by performing remedying activities such as studying feedback, and particularly when discrepancies between answers and solutions are large (Kulhavy & Stock, 1989). Productive failure provides opportunities for learners to study canonical instructions after failing during initial assessments, and discrepancies between initial answers and solutions are likely to be larger in PF environments than for most DI environments.

The notion that failure can be effective and even essential for learning is also supported by the imperfect mental model; according to this model, learners build imperfect mental models during the course of learning, and in order for repairs to be made to those models, learners must be able to recognize flaws in them (Chi, 2000). The initial assessments in PF provide opportunities for learners to realize flaws in understanding, while learners using DI are perhaps less likely to recognize flaws because merely applying canonical solutions to problem-solving is less likely to result in the types of failure that cause such realizations. These realizations in PF drive students to allocate mental resources on only the content most relevant to repairing their specific gaps in understanding (Durkin & Rittle-Johnson, 2012).

Furthermore, failures (plus ensuing canonical instruction) not only assist learners in recognizing flaws, they help learners identify reasons why a canonical solution works and why a non-canonical solution does not always work, which improves learners' abilities to transfer information to novel situations that do not involve the canonical solution (Kapur & Lee, 2009). One example is a study in which students who were allowed to enter incorrect formulas and to observe the consequences of doing so (as opposed to being corrected immediately upon the incorrect entry) indeed performed better on retention and transfer tasks than the immediately-corrected students (Mathan & Koedinger, 2003). These types of comparisons between generated solutions and canonical solutions enable learners to better encode critical conceptual features and therefore select correct problem-solving procedures, even in novel situations (Siegler, 2002). PF methods provide ample failure-driven comparison opportunities via the "problem-solving prior" order of instruction as well as the ill-structured nature of the initial problems, while DI methods are more conducive to learners simply "regurgitating" information that only applies to isomorphic problems. In short, failures are a necessary component of learning, especially for expertise and learning of deeper varieties, and PF appears to be well-suited for eliciting those failures.

3.1.3 Immediate performance vs. enduring learning. Even if tasks are not complicated enough to always produce failure, difficulties in tasks can still produce the desired effect of enduring learning in the face of decreased immediate performance; these difficulties are called "desirable difficulties" (Bjork, 2013). Desirable difficulties require "encoding and/or retrieval activities that support learning" (Bjork, 2013, p. 3), and a few examples of them are:

- *Environmental factors*: Using clutter items in a luggage screening training environment can improve test performance relative to clutter-absent training (Fiore, Scielzo, Jentsch, & Howard, 2006).
- *Training variation*: Kids who practice throwing beanbags to targets of varying distances (not including the tested distance) perform better on tests than those who practice at exactly and only the tested distance (Kerr & Booth, 1978).
- *Scheduling*: In learning environments with multiple tasks, interleaved practice scheduling (random determination of the task to be practiced next) yields slower improvement, during training, than blocked practice scheduling (practicing one task many times before switching to another task), but produces better retention performance, especially on interleaved assessments (e.g., Shea & Morgan, 1979).
- *Secondary tasks*: In a radar detection task, an irrelevant concurrent secondary task during training decreased primary task test performance, even when the test also

involved the irrelevant secondary task; however, adding another concurrent secondary task during training (this one being a relevant one, in addition to the primary and irrelevant secondary tasks) increased test performance, even when the test only involved the primary task and the irrelevant secondary task (Young et al., 2011).

It should be noted that any difficulties designed into learning must be relevant to the task in order to be "desirable" (i.e., produce learning gains; Young et al., 2011).

The notion that short-term performance is not correlated with long-term and/or transfer performance (e.g., Schmidt & Bjork, 1992; Goode & Magill, 1986) is central to questions about the effectiveness of PF methods. After all, failures in PF will almost always lead to relatively poor performances if learners are assessed soon after those failures, when compared to DI learners spending the same amount of time to that point. Instead, Soderstrom and Bjork (2015) argue that learning should be the goal of instruction ("permanent changes in comprehension, understanding, and skills of the types that will support long-term retention and transfer," p. 176), not immediate performance (possibly temporary measures that can be unreliable indicators of learning). Learning is a function of storage strength, an indication of how entrenched and associated new information is with related and prior knowledge, while immediate performance is a function of retrieval strength, which is often only high in particular contexts and when the material happens to be currently activated (Bjork & Bjork, 1992). Furthermore, Bjork (2013) adds that information cannot be learned by students making a "literal copy of that information" (p. 2), yet many DI methods appear to attempt exactly that, and thus many students have internalized activities like obsessive re-reading as legitimate material review strategies (84% of students use re-reading as a study strategy; Karpicke, Butler, & Roediger, 2009). To increase the information storage strength that is vital for enduring learning, students should retrieve information rather than review it because retrieval attempts help create “new routes” to information (testing effect; Carrier & Pashler, 1992) that activate various pre-existing knowledge along the way. This notion that test-like items can serve as learning events (instead of just assessments) corroborates the philosophies of methods such as PF; however, just 11% of students report using retrieval practice in their studying routines (Karpicke, Butler, & Roediger, 2009), implying that PF-related methods are not as favored with students.

Due to the fact that learning and immediate performance are apparently very different, it follows logically that learning could happen with no outward performance improvement associated with it, and also that outward performance could improve with no associated learning (Soderstrom & Bjork, 2015). Empirically, this observation can be seen in work ranging from early research on rats in mazes (rats that wander aimlessly in a maze still often demonstrate improved ability to navigate the maze later; Blodgett, 1929) to students learning descriptive statistics (students that did not generate correct solutions during initial “invention periods” outperformed DI students later, as long as conventional instruction was available after invention; Schwartz & Martin, 2004). But not only is immediate performance un-correlated with long-term learning and transfer ability, some methods that aim to improve immediate performance can actively undermine students’ attempts to achieve enduring learning. For example, most DI methods provide very frequent and/or specific feedback to learners in an effort to boost performance on similar problems. However, immediate feedback can sometimes turn into a crutch that shields learners from having to create high-level problem-solving strategies, while those that are forced to wander a bit without feedback do demonstrate such strategies (Cope & Simmons, 1994). Indeed, a meta-analysis of feedback timing reveals that learners using delayed feedback score 0.44 standard deviations higher on retention tests than those using immediate feedback (Kulik & Kulik, 1988). Being too specific with feedback could also cause learners to perform less investigation of their incorrect responses and therefore less self-assessment, perhaps because they are placated by the short-term performance increases (Goodman, Wood, & Hendrickx, 2004). Therefore, it is advantageous to withhold instruction that might increase short-term performance at the expense of durable learning.

3.2 Early results from productive failure research (a meta-analysis)

While PF research is still relatively young, enough has been done for a meta-analysis to be warranted. Eighteen papers were included in this meta-analysis, each of which either A) implemented PF methods directly, or B) mentioned PF as the inspiration for the design of the instruction in the experimental condition. Papers were not included if the authors posited after the fact that PF-related processes might have been involved (i.e., PF was not used to directly inform the experimental manipulation). Table 2 provides a summary of the findings from each study and the relevant specifications of each study.

Table 2.

Summary of studies that directly implemented productive failure manipulations

Authors	Domain	Scaffolding	Comparison group	Group size	Assessment	Result	Effect size (Cohen's d)
Glogger et al., 2013	Strategy evaluation	Prompts	Direct instruction	Individual	Conceptual	PF < DI	-0.75
Holmes et al., 2014	Math	Prompts	No scaffolding (N)	Team	Procedural, near transfer	PF = N	0
Holmes et al., 2014	Math	Prompts	No scaffolding (N)	Team	Conceptual	PF = N	0
Holmes et al., 2014	Math	Prompts	No scaffolding (N)	Team	Far transfer	PF > N	+0.1
Hsu, Kalyuga, & Sweller, 2015	Physics	None	Direct instruction	Individual	Immediate test	PF < DI	-0.06
Hsu, Kalyuga, & Sweller, 2015	Physics	None	Direct instruction	Individual	Delayed test	PF < DI	-0.04
Hsu, Kalyuga, & Sweller, 2015	Physics	Principles	Direct instruction	Individual	Immediate test	PF < DI	-0.06
Hsu, Kalyuga, & Sweller, 2015	Physics	Principles	Direct instruction	Individual	Delayed test	PF < DI	-0.04
Kapur & Bielaczyc, 2011 (Exp. 1)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF = DI	0
Kapur & Bielaczyc, 2011 (Exp. 1)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF > DI	+0.18
Kapur & Bielaczyc, 2011 (Exp. 1)	Math	None	Direct instruction	Individual	Conceptual	PF > DI	+0.65
Kapur & Bielaczyc, 2011 (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Procedural, near transfer	PF = VF	0
Kapur & Bielaczyc, 2011 (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Procedural, near transfer	PF = VF	0
Kapur & Bielaczyc, 2011 (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Conceptual	PF > VF	+0.39
Kapur & Bielaczyc, 2011 (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Far transfer	PF > VF	+0.16
Kapur & Bielaczyc, 2011 (Exp. 3)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF = DI	0
Kapur & Bielaczyc, 2011 (Exp. 3)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF = DI	0
Kapur & Bielaczyc, 2011 (Exp. 3)	Math	None	Direct instruction	Individual	Conceptual	PF > DI	+0.26
Kapur & Bielaczyc, 2011 (Exp. 3)	Math	None	Direct instruction	Individual	Far transfer	PF = DI	0

Authors	Domain	Scaffolding	Comparison group	Group size	Assessment	Result	Effect size (Cohen's d)
Kapur & Bielaczyc, 2012 (Exp. 1)	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.32
Kapur & Bielaczyc, 2012 (Exp. 1)	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.14
Kapur & Bielaczyc, 2012 (Exp. 2)	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.12
Kapur & Bielaczyc, 2012 (Exp. 2)	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.1
Kapur & Bielaczyc, 2012 (Exp. 3)	Math	None	Direct instruction	Team	Procedural, near transfer	PF = DI	0
Kapur & Bielaczyc, 2012 (Exp. 3)	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.1
Kapur, Dickson, & Yhing, 2010	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.42
Kapur, Dickson, & Yhing, 2010	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.98
Kapur & Kinzer, 2009	Physics	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.22
Kapur & Kinzer, 2009	Physics	None	Direct instruction	Team	Far transfer	PF > DI	+0.2
Kapur & Lee, 2009 (Exp. 1)	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.12
Kapur & Lee, 2009 (Exp. 1)	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.24
Kapur & Lee, 2009 (Exp. 2)	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.18
Kapur & Lee, 2009 (Exp. 2)	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.52
Kapur, 2011	Math	None	Direct instruction	Team	Procedural, near transfer	PF > DI	+0.2
Kapur, 2011	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.01
Kapur, 2012	Math	None	Direct instruction	Team	Procedural, near transfer	PF = DI	0
Kapur, 2012	Math	None	Direct instruction	Team	Conceptual	PF > DI	+0.98
Kapur, 2012	Math	None	Direct instruction	Team	Far transfer	PF > DI	+0.47

Authors	**Domain**	**Scaffolding**	**Comparison group**	**Group size**	**Assessment**	**Result**	**Effect size (Cohen's d)**
Kapur, 2014a (Exp. 1)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF = DI	0
Kapur, 2014a (Exp. 1)	Math	None	Direct instruction	Individual	Conceptual	PF > DI	+2.00
Kapur, 2014a (Exp. 1)	Math	None	Direct instruction	Individual	Far transfer	PF > DI	+1.52
Kapur, 2014a (Exp. 2)	Math	None	Direct instruction	Individual	Procedural, near transfer	PF = DI	0
Kapur, 2014a (Exp. 2)	Math	None	Direct instruction	Individual	Conceptual	PF > DI	+2.25
Kapur, 2014a (Exp. 2)	Math	None	Direct instruction	Individual	Far transfer	PF > DI	+1.29
Kapur, 2014a (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Procedural, near transfer	PF = VF	0
Kapur, 2014a (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Conceptual	PF > VF	+1.35
Kapur, 2014a (Exp. 2)	Math	None	Vicarious failure (VF)	Individual	Far transfer	PF > VF	+1.23
Kapur, 2014b	Math	Motivation	Vicarious failure (VF)	Team	Procedural, near transfer	PF > VF	+0.32
Kapur, 2014b	Math	Motivation	Vicarious failure (VF)	Team	Conceptual	PF > VF	+0.91
Kapur, 2014b	Math	Motivation	Vicarious failure (VF)	Team	Far transfer	PF > VF	+0.75
Loibl & Rummel, 2014a (Exp. 1)	Math	None	Direct instruction	Team	Procedural, near transfer	PF < DI	-0.07
Loibl & Rummel, 2014a (Exp. 1)	Math	None	Direct instruction	Team	Conceptual	PF > DI	+0.63
Loibl & Rummel, 2014a (Exp. 1)	Math	Contrasting cases	Direct instruction	Team	Procedural, near transfer	PF < DI	-0.07
Loibl & Rummel, 2014a (Exp. 1)	Math	Contrasting cases	Direct instruction	Team	Conceptual	PF > DI	+0.63
Loibl & Rummel, 2014a (Exp. 2)	Math	None	Direct instruction	Team	Procedural, near transfer	PF = DI	0
Loibl & Rummel, 2014a (Exp. 2)	Math	None	Direct instruction	Team	Conceptual	PF > DI	+0.63
Loibl & Rummel, 2014a (Exp. 2)	Math	Contrasting cases	Direct instruction	Team	Procedural, near transfer	PF = DI	0
Loibl & Rummel, 2014a (Exp. 2)	Math	Contrasting cases	Direct instruction	Team	Conceptual	PF > DI	+0.63

Authors	**Domain**	**Scaffolding**	**Comparison group**	**Group size**	**Assessment**	**Result**	**Effect size (Cohen's d)**
Loibl & Rummel, 2014b (Exp. 1)	Math	None	Direct instruction	Team	Procedural, near transfer	PF < DI	-0.18
Loibl & Rummel, 2014b (Exp. 1)	Math	None	Direct instruction	Team	Conceptual	PF > DI	+0.54
Loibl & Rummel, 2014b (Exp. 2)	Math	None	Direct instruction	Team	Procedural, near transfer	PF = DI	0
Loibl & Rummel, 2014b (Exp. 2)	Math	None	Direct instruction	Team	Conceptual	PF > DI	+0.16
Mazziotti et al., 2015 (Exp. 1)	Math	None	Direct instruction	Individual	General	PF > DI	+0.18
Mazziotti et al., 2015 (Exp. 1)	Math	Interaction scripts	Direct instruction	Team	General	PF > DI	+0.18
Mazziotti et al., 2015 (Exp. 2)	Math	None	Direct instruction	Individual	General	PF = DI	0
Mazziotti et al., 2015 (Exp. 2)	Math	Interaction scripts	Direct instruction	Team	General	PF = DI	0
Pathak et al., 2008	Physics	None	Direct instruction	Team	General	PF > DI	+0.33
Roll et al., 2013	Physics	Metacognition	No scaffolding (N)	Team	Procedural, near transfer	PF = N	0
Roll et al., 2013	Physics	Metacognition	No scaffolding (N)	Team	Conceptual	PF = N	0
Roll et al., 2013	Physics	Metacognition	No scaffolding (N)	Team	Far transfer	PF > N	Not found
Westermann & Rummel, 2012	Math	Interaction scripts	Direct instruction	Team	General	PF > DI	+0.67

From these studies, some preliminary outcomes can be observed regarding productive failure and, more generally, learner assistance. The sections below summarize some of the more notable outcomes from these studies in terms of the average performance differences between PF and comparison groups (in standard deviations).

3.2.1 Type of learning tested. Below are the results regarding how PF fared against comparison groups for three general types of learning:

- Near transfer, procedural: PF outperforms comparison groups by 0.06 SD
- Conceptual: PF outperforms comparison groups by 0.66 SD
- Far transfer: PF outperforms comparison groups by 0.49 SD

These results largely align with previously-mentioned hypotheses: The comparison groups (mostly DI-related) are comparable to PF in terms of helping students achieve near-transfer and procedural learning because of the similarity of the problems to the learning materials, while PF is much more viable in the deep learning of structural concepts and transferring to novel problems. Given that PF is a minimal-guidance method by design, it is unsurprising that students using PF performed better in ways usually associated with MG methods (e.g., generalizable knowledge). For example, PF learners developed complex and varied methods for exploring problem spaces that were used when solving ill-structured problems (Kapur & Kinzer, 2009). However, the fact that PF students also performed comparably well in near-transfer and procedural knowledge, realms in which direct instruction is usually more suitable, is perhaps a bit more interesting and indicates that achieving “the best of both worlds” (MG and DI) is possible. One observation that could help explain this phenomenon is that PF participants generally reported greater curiosity during the canonical instruction periods than those in DI conditions, likely because they had identified knowledge gaps during problem-solving and initial assessment periods and were interested in resolving those gaps (Loibl & Rummel, 2014b); in DI, participants generally receive canonical instruction without as inherent a reason to pay attention. If DI methods rely chiefly on canonical instruction as the vehicle for teaching, and PF methods are able to induce higher learner engagement with canonical instruction, then it follows that PF might be able to achieve many of the positive effects usually associated with DI.

3.2.2 Scaffolding levels. Some PF research has investigated whether scaffolding learners during the first problem-solving period could be beneficial for learners and the results are below:

- Scaffolded PF outperforms comparison groups by 0.18 SD
- Un-scaffolded PF outperforms comparison groups by 0.36 SD

Admittedly, the realm of scaffolding-PF interactions is not close to being fully investigated, especially given the many methods of scaffolding that could be implemented (but have not been investigated yet). Nevertheless, the preliminary finding that adding scaffolding to PF does not necessarily improve learning is logical, to some degree. Loibl and Rummel (2014a) write that adding scaffolding fundamentally changes the cognitive processes normally associated with productive failure: PF methods should not necessarily lead learners to discover canonical solutions, but induce the activation of prior knowledge and intuition for diverse solution approaches (Kapur & Bielaczyc, 2012); however, guiding learners with scaffolding can serve to reduce the problem space enough that they discover canonical solutions too easily, thereby depriving learners the benefits of productive failures (an issue with many MG methods). Therefore, key features in initial PF problem-solving should not be made salient (e.g., “contrasting cases” scaffolding methods might help learners guess canonical solutions too easily; Roll et al., 2009), although support for metacognition and learner interactions are examples of scaffolding that can encourage learners to build from prior knowledge and intuition without reducing the problem space and making key features too available (Loibl & Rummel, 2014a). Given that evidence exists demonstrating the effectiveness of guidance to some extent, much relevant work remains to be done in regards to the implementations of various scaffolding methods in PF.

3.2.3 Vicarious failure. An intriguing extension of the productive failure method is vicarious failure (VF), in which students learn from the failed solutions of their peers during the problem-solving period, as opposed to generating their own solutions and learning from their own failures. Given that diversity of invented solutions is predictive of learning, allowing students to study a variety of their peers' different invented solutions instead could confer some benefits that generating one's own (perhaps limited) set of solutions cannot (Kapur, 2014b). In the cognitive realm, it has been hypothesized that VF might induce lower cognitive load than PF, leaving more resources available for encoding of information and schema acquisition, thereby improving learning (Paas, Renkl, & Sweller, 2003). More specifically, generation-related activities are more burdensome to learners' existing domain knowledge than evaluation-related activities (Kapur, 2014a), implying that VF could be more effective for a broad range of students (as opposed to PF, which could require learners to have more background in the area).

However, those generation activities help students understand their invented solutions in ways that merely studying other students' failed solutions cannot; their understanding enables them to make deeper and more relevant comparisons to canonical solutions on critical features of problems (Terwel, van Oers, van Dijk, & van den Eeden, 2009). As a result, PF students are more likely to use correct procedures on future tasks, and according to the (relatively limited number of) VF-related studies in this meta-analysis, PF students indeed outperformed VF students by an average 0.51 standard deviations across all types of assessments, indicating that the structural and transferable understanding associated with PF is "worth" the high cognitive load. One of these studies found that PF students reported higher mental effort than VF students (corroborating the cognitive load concerns), but also scored more highly on "conceptual understanding and transfer without compromising procedural fluency" (Kapur, 2014b, p. 651), a finding that diverges from what would be predicted by cognitive load theory. Perhaps a "sweet spot" of cognitive load does exist in regards to performance (e.g., Lens, 1983; Vygotsky, 1978), with VF-induced load not being quite high enough, or maybe inventing solutions has positive motivating effects on students (Belenky & Nokes-Malach, 2012). In any case, the cognitive processes involved in inventing solutions and learning from one's own mistakes appear to improve learning over mere evaluation of failed solutions, but it remains to be seen whether failure itself is necessary or whether characteristics such as the blending of formal and intuitive knowledge are most important (Kapur & Rummel, 2012).

3.2.4 Group size. The meta-analysis revealed these data regarding how students learn in PF individually and in groups:

- Individual learning: PF outperforms comparison groups by 0.39 SD
- Team learning: PF outperforms comparison groups by 0.27 SD

Given that collaborative learning often elicits elaborative cognitive processes (e.g., Teasley, 1995), and that to some extent, the positive effects of PF are driven by the diversity of invented solutions, a reasonable hypothesis would be that teams using PF should generally perform better than individuals. Intuitively, team approaches to PF could also be beneficial because any given set of multiple opinions is more likely to "normalize" (i.e., be reasonably close to the canonical solution) and therefore avoid the sort of outlandishness that could steer a learner to unproductive failure.

However, the data indicate that adding team members does not appear to improve the effectiveness of PF over working individually. One explanation for this finding involves the notion that learner groups often need to be externally supported in order for collaborations to be successful (King, 2007): The design of PF is such that learners are unsupported during initial problem-solving periods, suggesting that team interactions might not have been very productive during those periods when compared to individual learners. Another hypothesis is that the turn-taking nature of collaborative solution generation limits the flow of invented solution ideas (Stroebe & Nijstad, 2004), which is particularly problematic for PF methods because diverse solutions are key; obviously, individual learners do not have this turn-taking issue and can instead generate ideas much more freely. Scaffolding methods related to facilitating collaboration have some demonstrated effectiveness and should be implemented in future team-learning PF research, although Mazziotti et al. (2015) note that learner familiarity with the collaboration support mechanisms could play a large role in the effectiveness of those mechanisms.

CHAPTER FOUR

Future directions in learner assistance research

Given that the notion of productive failure is still in its early stages, much work remains to be done in terms of exploring unresolved PF issues and, in so doing, revealing more about the fundamental nature of learner assistance.

4.1 Current issues with PF

Although productive failure appears to be promising in producing durable and generalizable learning, some problems persist that could require further investigation of the phenomenon.

4.1.1 Ownership of learning. Educators often state that they want students to take ownership of the learning process, implying that when students are empowered in that way, they will become more self-motivated and feel more responsible for their actions, thereby enhancing learning. It undoubtedly brings some benefits to learners; for example, Kapur (2014b) hypothesizes that generating solution methods engenders feelings of ownership, which in turn causes learners to be more interested in learning canonical solutions. However, ownership of the learning process could create some unexpected consequences that give pause to researchers. These consequences might be especially pronounced for students in PF environments who, early on in the learning process, will likely have more misconceptions than their DI counterparts (Brown & Campione, 1994). The continued influence effect states that incorrect information can persist in influencing students' understanding of materials even after the incorrect information has been discredited (e.g., Johnson & Seifert, 1994), and to compound the problem, students can also be influenced by the "IKEA effect," the tendency for people to generally value self-created items over others' items because the process of creation fulfills a psychological need for people to feel competent (Mochon, Norton, & Ariely, 2012).

To some extent, PF environments are particularly susceptible to both of these effects. It is easy to imagine situations in which students invent flawed solution methods during Period 1, and then continue to be negatively influenced by them (despite correction periods later in the learning process) because they feel a sense of ownership regarding those methods and subconsciously prefer them over the canonical methods designed by instructors. However, the continued influence effect of incorrect information can be largely mitigated through the use of discrediting information that A) provides plausible alternative explanations for learned concepts, and B) clarifies any outstanding pieces of information that might arise from the negation of an incorrect concept (Johnson & Seifert, 1994). Therefore, the canonical instruction in PF correction periods should strive not only to negate previously-held misconceptions in students (if only negation is done, students might make more erroneous inferences to tie up loose ends and outstanding information to create a whole understanding; Johnson & Seifert, 1994), but provide solutions and convince students that those solutions are more plausible than the ones they created.

4.1.2 Optimal amounts of struggle. While early results from PF research are encouraging, it is not generally known whether the failures induced in PF are as productive as they could be. Perhaps a more-optimal amount of learner struggle exists that is still yet to be discovered. Some research has indicated that the number of high-quality solutions invented during generation periods (the initial problem-solving periods) is a better predictor of learning outcomes than the number of low-quality solutions invented (Wiedmann et al., 2012); therefore, while struggle is important (i.e., not being able to figure out the canonical solution), the work produced through the struggle needs to be sufficiently on-track to be productive. After all, misconceptions from off-track learning can be hard to shake: For example, misconceptions about the human heart learned in elementary school can sometimes linger until the students attend college and even beyond (Ozgur, 2013). Wiedmann and colleagues (2012) also add that when learning in groups, the presence of at least one member of high ability supports discussion and helps the group create more solution methods to therefore learn more from the inventing process. All of these findings substantiate the notion that the struggles associated with failures are not necessarily productive if they merely amount to learners "flailing" aimlessly – struggles should be tough enough to elicit the beneficial failure-related cognition discussed earlier in this document, but forgiving enough that students still learn substantive information.

In fact, failure itself might not even be always necessary for enduring learning. To some extent, students being warned of likely errors can improve learning in a way that avoids the need for learners to commit all of the errors themselves (Loibl & Rummel, 2014b). Perhaps the most important element of instruction is that students perform actions that are commonly associated with failure-related methods (e.g., creating a diversity of solution methods, undergoing expectation failures, recognizing knowledge gaps, learning generalizable methods, becoming curious and motivated to learn), with the failures themselves being ancillary. As of now, it appears that failures are the most effective way to elicit these actions, but if non-failure methods arise that can achieve the same benefits, they might be more useful because of inherent advantages regarding learner preferences: Some evidence indicates that learners generally dislike minimally-guided methods such as PF and subjectively rate material as difficult when it is presented using a MG method, likely due to excessive cognitive load (Reynolds & Caperton, 2011). Interestingly, although high-intelligence learners tend to perform better with less guidance, they tend to prefer heavily-guided versions of courses because of the minimal effort they believe will be necessary to complete them (Clark, 1982). This information about preference is important because the implementation of a method will encounter less resistance if students are more amenable to it.

4.1.3 Hard-easy effect. In the debate between minimal guidance and direct instruction, the literature presents conflicting data concerning the methods that subject learners to lower cognitive load. DI methods tend to use reductions in problem-space searching as the chief mechanism for load reduction (e.g., Kirschner et al., 2006) while MG methods tend to rely on greater activation of prior knowledge (e.g., Kapur & Bielaczyc, 2011). However, for PF in particular, the evidence appears to point toward the fact that the method might indeed induce higher cognitive load, whether in comparison to vicarious failure (VF) or direct instruction (Kapur, 2014a).

This higher load in PF, in conjunction with its relatively high subjective difficulty ratings (as discussed earlier), could have implications beyond just learning and performance. The hard-easy effect states that learners studying difficult materials tend to be overconfident in their understanding relative to when they study easier materials (Lichtenstein, Fischhoff, & Phillips, 1982). Therefore, PF is more likely than any given DI method to cloud processes related to evaluating one's own learning, which in turn might also affect learning and performance. Perhaps metacognitive support such as reflection prompts (van den Boom, Paas, van Merrienboer, & van Gog, 2004) during initial problem-solving, the period most likely to induce high cognitive load, could be useful in controlling students' overconfidence.

4.2 Potential future directions in PF

Future investigations of productive failure could proceed in many different directions. However, two important and potentially intriguing directions have been particularly under-investigated, and they are outlined in the following sections.

4.2.1 Scaffolding during PF generation periods. The meta-analysis presented in this document demonstrates that productive failure can be more effective than other minimal-guidance and direct instruction methods. However, those results do not imply that current PF methods are necessarily optimal. Many of the studies did not implement scaffolding during initial problem-solving, likely because the researchers wanted to ensure failures of some sort during the generation periods. While that objective is pedagogically meaningful, it should also be noted that guidance in some form is likely beneficial for discovery-based methods (e.g., Anthony, 1973). Furthermore, the studies that did implement scaffolding did not cover many of the possible scaffolding methods that could be used. Therefore, many scaffolding-PF interactions remain to be researched, some of which might produce effects better than non-scaffolded PF methods would alone, especially those scaffolding methods that provide just enough guidance to ensure that failures are productive. Two such examples of these scaffolding methods, which were examined earlier in this document, are discussed here.

- *Subgoals* (e.g., Catrambone, 1998): One of the ways in which learners might fail unproductively is by misunderstanding deep structures of the problem space; for example, being confused about the various objectives required to solve a problem could cause learners to take actions that are immaterial and contribute little to learning the task at hand, wasting time and energy as well as inducing frustration and perhaps unproductive failures (false starts; Carlson, Lundy, & Schneider, 1992). The use of subgoals could ensure that learners at least recognize the fundamental objectives associated with a problem, therefore increasing the likelihood that any learner actions and associated failures are at least correct in their intentions. Furthermore, because subgoals merely inform learners about functional groupings of steps, as opposed to directly instructing the mechanics of solving problems, learners are still likely to fail (productively) in generating canonical solutions. Of course, a "sweet spot" likely exists regarding the level of guidance provided by the subgoals (e.g., granularity of subgoals).

- *Training wheels* (e.g., Carroll & Carrithers, 1984): Another way to ensure that learner failures are productive is by explicitly locking learners out of areas of the problem space that are unlikely to reveal useful information. Whereas with subgoals, learners are instructed such that they have correct intentions in mind (but are not necessarily stopped from going off-track), training wheels provide nudges when learners do veer too far from canonical solutions, thereby preventing them from working unproductively. In the realm of learning how to use a word processor, Carroll and Carrithers (1984) prevented learners from reaching error states that would have produced more frustration than learning. Similar processes could be used in more traditionally academic domains (e.g., informing students that solving a given physics problem does not involve the use of a particular constant). If designed properly, the training wheels approach could allow learners a certain amount of freedom to explore and fail, but limit learners to failures that are likely to be productive; assuming that the training wheels are not designed such that it restricts learner actions to just canonical solutions, this method could be promising if used in conjunction with PF.

4.2.2 Domains of study. As of now, almost all PF research has been conducted in the domains of physics and math, which are traditional academic subjects and involve relatively closed-ended problems. There are reasons to believe, however, that the effectiveness of PF might change based on the domain of study. For domains in which rote memorization is key to success, such as introductory biology (many vocabulary terms) or history (names and dates), PF might not be a useful method because transfer ability is a relatively small part of learning the domain. That is, one of the key benefits of productive failures (generalizable knowledge) would be less important in these domains; the cognitive costs of failing might no longer be worth the remaining benefits. However, for domains in which creativity and novel problems are inherent such as musical composition, the diverse solution methods (diSessa & Sherin, 2000) and deeper informational connections (Kapur & Bielaczyc, 2011) generated by PF will likely be of great utility because acceptable solutions will vary widely. After all, it would be unsurprising if two students arrived at the same understanding of an elementary biology concept, but it is not likely that any two given students compose the same short song, even if instructed about the same general concepts to be used in the song. Collins (2012) adds that “most problems in life do not come with a canonical solution” (p. 734) and posits that PF might be “a more powerful way to teach” open-ended domains such as critical reading.

An even further extension of studying different domains is to move beyond academic domains and into group tasks, in which students learn to work together on problems/tasks that are team-oriented in nature. Some existing PF research has been carried out with groups of students, but the domains (again, mostly physics and math) required individualized understanding and did not involve team situation awareness or anything related to a shared group understanding. With team-oriented tasks, two types of training methods might be able to induce productive failures during initial problem-solving (generation periods):

- *Cross-training*: People learn the responsibilities of other group members to develop a shared knowledge; this learning can take place through “positional clarification (receiving information on other roles), positional modeling (observing other roles), and positional rotation (firsthand experience performing different roles)” (Blickensderfer, Cannon-Bowers, & Salas, 1998). When learning about the other roles in the group, team members will likely stumble because their tasks might be very different from their original trained tasks. However, the failures associated with performing these other roles should help reveal fundamental information about how the team should interact. Cross-training has been shown to improve performance by helping team members anticipate each other’s needs (e.g., Stout, Cannon-Bowers, Salas, & Milanovich, 1999) and predict appropriate actions for novel situations (Fiore, Salas, & Cannon-Bowers, 2001). During initial problem-solving in PF, cross-training will almost definitely decrease team performance in the short-term, but should produce generalizable knowledge that improves performance in the long run and in various contexts.
- *Perturbation training*: Perturbations are extrinsic disruptions to team processes that force teams to coordinate in new ways to achieve objectives; the goal of this style of training is to “counteract habituation and procedural rigidity associated with team interactions…allowing teams to acquire flexible interaction processes” (Gorman, Cooke, & Amazeen, 2010, p. 297). If severe enough, these disruptions

could induce failure to initially achieve objectives and canonical solutions, which could help learners by breaking knowledge stability biases and overconfidence (Kornell & Bjork, 2009), encouraging them to create diverse solutions sets (Clement, 1991; Reif & Larkin, 1991), and varying practice conditions such that the practiced cognitive processes are more like those needed in post-training environments (Bjork, 1994). Perturbation training has been demonstrated to increase performance in novel conditions (Gorman, Cooke, & Amazeen, 2010), and is therefore a good fit for use during initial problem-solving in PF environments.

CHAPTER FIVE

Conclusions

For all of its merits as an instructional method (durability of learning, transferability of learning to various contexts, deep structural understanding of domain, high learner engagement, etc.), perhaps one of the most important contributions of productive failure (PF) research is what it reveals about the fundamental nature of learner assistance. The two most important revelations are outlined below:

- Learner struggle (and explicit failure) can be leveraged to produce very powerful learning outcomes, which are driven by cognitive processes that are not elicited as frequently when learners do not struggle or fail. Minimal guidance (MG) and direct instruction (DI) methods alike tend to avoid struggle/failure as learning mechanisms.
- The learner benefits associated with MG and DI are usually thought to be specific to their respective methods, but achieving all of them with one method is possible (e.g., productive failure). Therefore, the learner assistance debate (often portrayed as "MG vs. DI" in the literature) as it stands now is perhaps not illustrated completely. Furthermore, both types of benefits are important for students to glean during the learning process.

As described previously, PF research is at a relatively early stage; therefore, much work remains to be done, not only in terms of validating the findings, but improving on the method as well. However, these findings, at a general level, can be used to improve existing instructional methods or create new ones, regardless of how those methods are classified. The findings also provide some more nuance to a debate that has been contentious at times and will likely lack a definitive solution.

References

Alexander, P.A. (2003). The Development of Expertise: The Journey From Acclimation to Proficiency. *Educational Researcher*, *32* (8), 10-14.

Anthony, W.S. (1973). Learning to discover rules by discovery. *Journal of Educational Psychology*, *64* (3), 325-328.

Aulls, M.W. (2002). The contributions of co-occurring forms of classroom discourse and academic activities to curriculum events and instruction. Journal of Educational Psychology, 94, 520–538.

Baddeley, A.D. & Longman, D.J.A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*, 627-635.

Barrows, H.S., & Tamblyn, R.M. (1980). Problem-based learning: An approach to medical education. New York: Springer.

Belenky, D.M. & Nokes-Malach, T.J. (2012). Motivation and Transfer: The role of Mastery-Approach Goals in Preparation for Future Learning. *Journal of the Learning Sciences, 21* (3), 399-432.

Bjork, E.L. & Bjork, R.A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher,et al. (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society*. New York: Worth Publishers.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R.A. (2013). Desirable difficulties perspective on learning. In H. Pashler (ed.), *Encyclopedia of the mind*. Thousand Oaks, CA: Sage Reference.

Bjork, R.A. & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes, Vol. 2* (pp. 35-67). Hillsdale, NJ: Erlbaum.

Blickensderfer, E., Cannon-Bowers, J.A., & Salas, E. (1998). Cross-training and team performance. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 299-311). Washington, DC: American Psychological Association.

Blodgett, H. C. (1929). The effect of the introduction of reward upon the maze performance of rats. *University of California Publications in Psychology*, *4*, 113–134.

Bransford, J.D. & Schwartz, D.L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P.D. Pearson (Eds.), *Review of research in education*. American Educational Research Association: Washington, DC, 61–101.

Brown, A. & Campione, J. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-270). Cambridge, MA: MIT Press.

Brown, J.S., Collins, A., & Duguid, P. (1988). Situated cognition and the culture of learning. *Educational Researcher*, *18* (1), 32-42.

Bruner, J.S. (1961). The art of discovery. *Harvard Educational Review, 31*, 21–32.

Carlson, R.A., Lundy, D.H., & Schneider, W. (1992). Strategy guidance and memory aiding in learning a problem-solving skill. *Human Factors, 34,* 129-145.

Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition, 20,* 633-642.

Carroll, J.M. & Carrithers, C. (1984). Training Wheels in a User Interface. *Communications of the ACM*, *27* (8), 800-806.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127* (4), 355-376.

Catrambone, R. & Carroll, J.M. (1987). Learning a word processing system with training wheels and guided exploration. *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface* (pp. 169-174). New York, NY: Association for Computing Machinery.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science, 5*, 121-152.

Chi, M.T.H. (2000). Self-explaining: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), Advances in instructional psychology, (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Clark, R.E., Kirschner, P.A., & Sweller, J. (2012). Putting students on the path to learning: the case for fully guided instruction. *American Educator*, *36*, 6-11.

Clement, J. (1991). Non-formal reasoning in science: The use of analogies, extreme cases, and physical intuition. In Voss, J. F., Perkins, D. N., & Siegel, J. (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. (1991). Cognitive apprenticeship and instructional technology. In L. Idol & B.F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 119-136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Collins, A. (2012). What is the most effective way to teach problem solving? A commentary on productive failure as a method of teaching. *Instructional Science*, *40*, 731-735.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining procedures and concepts. *Journal of Experimental Psychology: Applied*, *7*, 68–82

Cope, P., & Simmons, M. (1994). Some effects of limited feedback on performance and problem-solving strategy in a logo microworld. *Journal of Educational Psychology*, *86* (3), 368-379.

D’Ydewalle, G. (1984). Motivational and information processing. Unpublished report, University of Leuven, Belgium.

De Leeuw, L. (1983). Teaching problem solving: An ATI study of the effects of teaching algorithmic and heuristic solutions methods. *Instructional Science*, *12*, 1-48.

DeCaro, M.S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113* (4), 552–568.

Dean, D. & Kuhn, D. (2006). Direct Instruction vs. Discovery: The Long View. *Science Education*, *91*, 384-397.

Deci, E.L. & Ryan, R.M. (2000). The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11* (4), 227-268.

Dillenbourg, P. (2002). Over-scripting CSCL. In P.A. Kirschner (Ed.), Three worlds of CSCL: Can we support CSCL (pp. 61–91). Heerlen: Open University of the Netherlands.

diSessa A.A. & Sherin, B.L. (2000). Meta-representation: an introduction. *Journal of Mathematical Behavior*, *19*, 385-398.

diSessa, A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Children's meta-representational expertise. *Journal of Mathematical Behavior*, *10* (2), 117-160.

Dunlosky, J. & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249-376). Mahwah, NJ: Erlbaum.

Dunning, D., Johnson, K., Erlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83-87.

Durkin, K. & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22* (3), 206-214.

Edelson, D.C. (2001). Learning-for-use: A framework for integrating content and process learning in the design of inquiry activities. *Journal of Research in Science Teaching, 38*, 355–385.

Elshout, J.J. (1985). Problem solving and education. Paper presented at the First Conference of the European Association for Research on Learning and Instruction, Leuven, Belgium

Entwistle, N.J. (1985). Explaining individual differences in school learning. Paper presented at the First Conference of the European Association for Research on Learning and Instruction, Leuven, Belgium.

Fiore, S.M., Salas, E., & Cannon-Bowers, J.A. (2001). Group dynamics and shared mental model development. In M. London (Ed.), *How people evaluate others in organizations* (pp. 309–336). Mahwah, NJ: Erlbaum.

Fiore, S.M., Scielzo, S., & Jentsch, F., & Howard, M.L. (2006). Understanding performance and cognitive efficiency when training for x-ray security screening. *Proceedings of the Human Factors and Ergonomics society 50th Annual Meeting* (pp. 2610-2614). Santa Monica, CA: Human Factors and Ergonomics Society.

Glogger, I., Kappich, J., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2013). Inventing prepares learning motivationally, but a worked-out solution enhances learning outcomes. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2416-2421). Austin, TX: Cognitive Science Society.

Goode, S. & Magill, R.A. (1986). The contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport. 57*, 308-314.

Goodman, J.S., Wood, R.E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology, 89* (2), 248-262.

Gorman, J.C., Cooke, N.J., & Amazeen, P.G. (2010). Training Adaptive Teams. *Human Factors*. *52*, 295-307.

Gray, L.E. (1982). Aptitude constructs, learning processes, and achievement. Unpublished report. Stanford University.

Guzdial, M. (1997, June 16). Constructivism vs. Constructivism vs. Constructionism. Retrieved from: http://guzdial.cc.gatech.edu/Commentary/construct.html

Hagtvet, K.A. (1986). Interaction of anxiety and ability on academic achievement: A simultaneous consideration of parameters. Paper presented at the American Educational Research Association, San Francisco.

Hardiman, P., Pollatsek, A., & Weil, A. (1986). Learning to understand the balance beam. *Cognition and Instruction*, *3*, 1–30.

Hiebert, J., & Grouws, D.A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.

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.

Holmes, N.G., Park, A.K., Day, J., Bonn, D.A., & Roll, I. (2014) Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, *42* (4), 523-538

Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology, 3*(1), 37-53.

Johnson, H.M. & Seifert, C.M. (1994). Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20* (6), 1420-1436.

Jonassen, D. (1991). Objectivism vs. constructivism. *Educational Technology Research and Development*, *39* (3), 5–14.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). Expertise reversal effect. Educational Psychologist, 38, 23–31.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26* (3), 379-424.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: unpacking the design components. *Instructional Science*, *39*, 561-579.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, *40* (4), 651–672.

Kapur, M. (2014a). Productive Failure in Learning Math. *Cognitive Science*, *38*, 1008-1022.

Kapur, M. (2014b). Comparing Learning from Productive Failure and Vicarious Failure. *Journal of the Learning Sciences*, *23* (4), 651-677.

Kapur, M. & Bielaczyc, K. (2011). Classroom-based experiments in productive failure. In L. Carlson, C. Holscher, & T. Shipley (Eds.), Proceedings of the 33rd annual conference of the Cognitive Science Society (pp. 2812–2817). Austin: Cognitive Science Society.

Kapur, M. & Bielaczyc, K. (2012). Designing for Productive Failure. *Journal of the Learning Sciences*, *21* (1), 45-83.

Kapur, M., Dickson, L., & Yhing, T.P. (2010). Productive failure in mathematical problem solving. *Instructional Science*, *38* (6), 523–550.

Kapur, M. & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning*, *4* (1), 21-46.

Kapur, M., & Lee, K. (2009). Designing for productive failure in mathematical problem solving. In *Proceedings of the 31st Annual Conference Of The Cognitive Science Society* (pp. 2632-7). Austin, TX: Cognitive Science Society.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471–479.

Kerr, R. & Booth, B. (1978). Specific and varied practice of a motor skill. *Perceptual and Motor Skills*, *46*, 395-401.

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), Scripting computer-supported collaborative learning. *Cognitive, computational, and educational perspectives* (pp. 13–37). New York: Springer.

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 Learning. *Educational Psychologist*, *41* (2), 75-86.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2* (2–3), 211–224.

Koedinger, K.R. & Aleven, C. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239-264.

Koedinger, K.R., Corbett, A.T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, *36*, 757-798.

Kollar, I., Fischer, F., & Slotta, J.D. (2005). Internal and external collaboration scripts in web-based science learning at schools. In T. Koschmann, D. Suthers, & T.W. Chan (Eds.), Computer supported collaborative learning 2005: The Next 10 Years (pp. 331–340). Mahwah, NJ: Erlbaum.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*, 449–468.

Kornell, N. & Metcalfe, J. Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609-622.

Kulhavy, R.W., & Stock, W.A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1* (4), 279-307.

Kulik, J.A., & Kulik, C.C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58* (1), 79-97.

Lens, W. (1983). Achievement motivation, test anxiety, and academic achievement. *Psychological Reports*, University of Leuven, Belgium.

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.

Lin, X., Hmelo, X., Kinzer, C.K., & Secules, T.J. (1999). Designing technology to support reflection. *Educational Technology, Research and Development*, *47* (3), 43-62.

Loibl, K. & Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, *42*, 305-326.

Loibl, K. & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74-85.

Machiels-Bongaerts, M., Schmidt, H.G., & Boshuizen, H.P.A. (1993). Effects of mobilizing prior knowledge on information processing: Studies of free recall and allocation of study time. *British Journal of Psychology*, *84* (4), 481-498.

Mathan, S. Koedinger, K.R (2003). Recasting the Feedback Debate: Benefits of Tutoring Error Detection and Correction Skills. In H.U. Hoppe et al. (Eds.), *Artificial Intelligence in Education*, (pp. 13-20). IOS Press.

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.

Mazziotti, C., Loibl, K. & Rummel, N. (2015). Collaborative or Individual Learning Within Productive Failure: Does the Social Form of Learning Make a Difference? *Computer Supported Collaborative Learning*.

McLaughlin, A.C., Rogers, W.A., & Fisk, A.D. (2006). How Effective Feedback for Training Depends on Learner Resources and Task Demands. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 2624-2628). Santa Monica, CA: Human Factors and Ergonomics Society.

McNamara D.S. (2001). Reading Both High-Coherence and Low-Coherence Texts: Effects of Text Sequence and Prior Knowledge. *Canadian Journal of Experimental Psychology*, *55* (1), 51-62.

McNamara, D.S. & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-287.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131* (3), 349-363.

Metcalfe, J. & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language*, *52*, 463-477.

Mochon, D., Norton, M.I., & Ariely, D. Bolstering and Restoring Feelings of Competence via the IKEA Effect. *International Journal of Research in Marketing*, *29* (4), 363-369.

Nussbaum, E.M., Hartley, K., Sinatra, G.M., Reynolds, R.E., & Bendixen, L.D. (2002, April). Enahancing the quality of on-line discussions. Paper presented at the Annual meeting of the American Educational Research Association.

Ozgur, S. (2013). The persistence of misconceptions about the human blood circulatory system among students in different grade levels. *International Journal of Environmental & Science Education*, *8* (2), 255-268.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38,* 1–4.

Patel, V.L., Groen, G.J., & Norman, G.R. (1993). Reasoning and instruction in medical curricula. *Cognition & Instruction*, *10*, 335–378.

Pathak, S.A., Jacobson, M.J., Kim, B., Zhang, B., & Feng, D. Learning the physics of electricity with agent-based models: The paradox of productive failure. *Proceedings of the 17th International Conference on Computers in Education* (pp. 221-228). Taipei, Taiwan: Asia-Pacific Society for Computers in Education.

Pea, R.D. (2004). The social and technological dimensions of "scaffolding" and related theoretical concepts for learning, education and human activity. *Journal of the Learning Sciences*, *13* (3), 423–451.

Peterson, P.L. (1987). Selecting students and services for compensatory education. In B.I. Williams, P.A. Richmond, & B.J. Mason (Eds.), *Designs for compensatory education: Conference proceedings and papers*. Washington, D.C.: Research and Evaluation Associates, Inc.

Reif, F. & Larkin, J. H. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. *Journal of Research in Science Teaching (Special Issue: Students' models and epistemologies), 28*(9), 733-760.

Reiser, B.J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*, 273-304.

Reynolds, R. & Caperton, I.H. (2011). Contrasts in student engagement, meaning-making, dislikes, and challenges in a discovery-based program of game design learning. *Educational Technology Research and Development*, *59*, 267-289.

Roll, I., Aleven, V., & Koedinger, K.R. (2009). Helping students know 'further' – increasing the flexibility of students' knowledge using symbolic invention tasks. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1169–1174). Austin, TX: Cognitive Science Society.

Roll, I., Holmes, N.G., Day, J., Park, A.H.K., & Bonn, D.A. (2013). Process and outcome benefits for orienting students to analyze and reflect on available data in productive failure activities. *Proceedings of the Workshop on Scaffolding in Open-Ended Learning Environments (OELE), in conjunction with Artificial Intelligence in Education 2013* (pp. 61-68). Memphis, TN.

Rourke, A. & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning and Instruction*, *19*, 185-199.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, *2* (3), 235-276.

Rutherford, F.J. (1964). The role of inquiry in science teaching. *Journal of Research in Science Teaching*, *2*, 80–84.

Schank, R. (1997). *Virtual Learning: A Revolutionary Approach to Building a Highly-Skilled Workforce*. New York, NY: McGraw-Hill

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.

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.

Shea, J.B. & Morgan, R.L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179-187.

Siegler, R.S. (2002). Microgenetic studies of self-explanation. In N. Garnott & J. Parziale (Eds.), *Microdevelopment: A process-oriented perspective for studying development and learning* (pp. 31–58). Cambridge, UK: Cambridge University Press.

Snow, R.E. (1982). Education and intelligence. In R.J. Sternberg (Ed.), *Handbook of Human Intelligence*. New York: Cambridge University Press.

Snow, R.E. (1989). Aptitude-Treatment Interaction as a framework for research on individual differences in learning. In P. Ackerman, R.J. Sternberg, & R. Glaser (eds.), *Learning and Individual Differences*. New York: W.H. Freeman.

Soderstrom, N.C. & Bjork, R.A. (2015). Learning versus Performance: An Integrative Review. *Perspectives on Psychological Science*, *10* (2), 176-199.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2* (4), 421-447.

Stout, R.J., Cannon-Bowers, J.A., Salas, E., & Milanovich, D.M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors*, *41*, 61–71.

Stroebe, W. & Nijstad, B.A. (2004). Warum Brainstorming in Gruppen Kreativität vermindert: Eine kognitive Theorie der Leistungsverluste beim Brainstorming [Why collaborative brainstorming reduces creativity: a cognitive theory on performances loses in brainstorming]. *Psychologische Rundschau*, *54*, 2-10.

Sullivan, A.M. & Skanes, G.R. (1971). Differential transfer of training in bright and dull subjects of the same mental age. *British Journal of Educational Psychology*, *41* (3), 287-293.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, *8* (4), 351–362.

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.

Sweller, J., Mawer, R., & Howe, W. (1982). The consequences of history-cued and means-ends strategies in problems solving. *American Journal of Psychology, 95,* 455–484.

Teasley, S.D. (1995). Communication and collaboration: The role of talk in children's peer interactions. *Developmental Psychology, 31* (2), 207-220.

Terwel, J., van Oers, B., van Dijk, I.M.A.W., & van den Eeden, P. (2009). Are representations to be provided or generated in primary mathematics education? Effects on transfer. *Educational Research and Evaluation, 15* (1), 25–44.

Toth, E.E., Suthers, D.D., & Lesgold, A.M. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education, 86* (2), 264-286.

van den Boom, G., Paas, F., van Merrienboer, J.J.G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: effects on students' self-regulated learning competence. *Computers in Human Behavior, 20,* 551-567.

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36,* 212-218.

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, 21* (3), 209-249.

Vygotsky, L. S. (1978). *Mind in society: The development of higher order psychological processes*. Cambridge, England: Cambridge University Press.

Westermann, K. & Rummel, N. (2012). Delaying instruction: evidence from a study in a university relearning setting. *Instructional Science, 40* (4), 673-689.

Wiedmann, M., Leach, R.C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40,* 711-730.

Winograd, T. & Flores, F. (1987). *Understanding Computers and Cognition: A New Foundation for Design*. Boston, MA: Addison-Wesley Publishing.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17,* 89-1 00.

Young, M.D., Healy, A.F., Gonzalez, C., Dutt, V., & Bourne, L.E. (2011). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*, *25*, 395-407.