The Effects of Experience Grouping on Achievement, Problem-Solving Discourse, and Satisfaction in Professional Technical Training

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Robert Sean Mulcahy

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Aaron Doering, Adviser
Charles Miller, Co-Adviser

August, 2010

Acknowledgements

This research would not have been possible without the help of:

- Vanessa Kubota, my chief subject matter expert, instructional design partner, and co-researcher.
- Danielle Dupuis of the University of Minnesota's Department of Educational Psychology, my consultant for matters statistical.
- The leadership of learning and professional development at my firm, particularly Jeanine Hansen, Dave Boizelle, Scott Johnson, and Jacque Anderson.
- Dr. Aaron Doering and Dr. Charles Miller, my advisers.

My sincere thanks to each of them. Above all, thank you to my wife Diana for her patience and support.

Abstract

Learners inevitably enter adult technical training classrooms—indeed, in all classrooms—with different levels of expertise on the subject matter. When the diversity of expertise is wide and the course makes use of small group problem solving, instructors have a choice about how to group learners: they may distribute learners with greater expertise among the groups so they can provide tutoring to their less experienced classmates, or instructors may choose to segregate learners to keep those with greater expertise from dominating problem solving. The literature base on ability grouping in small groups, the bulk of which has been conducted on K-12 learners, provides no one-size-fits-all solution, and no systematic research has been conducted on post-college adult learners in technical training situations nor on learners grouped together to solve relatively ill-structured problems. Therefore, the purpose of this research is to study the relationship between experience grouping and learning gains in small groups of adults learning to solve technical, relatively ill-structured problems.

Forty-six newly hired auditors beginning their careers by taking a mandatory course centered around relatively ill-structured problem solving were assigned to stratified random groups by experience level so that some groups featured similar levels of experience and other groups featured a relatively wide range of experience. No significant differences in learning gains or satisfaction were detected between heterogeneous and homogeneous groups. Analysis of qualitative data, however, indicated that small groups exhibited significantly different behaviors depending on whether they were asked to solve relatively well-structured or ill-structured problems. When groups were asked to solve relatively ill-structured problems, learners were three to five times as likely to exhibit behaviors associated with learning in problem-solving groups. Further, analysis of the qualitative data suggests with marginal significance that during relatively ill-structured problem solving there is a strong parabolic relationship between the spread of experience in the small groups and the percentage of positive collaborative problem-solving strategies: as experience spread rose, groups either sought a larger or smaller

percentage of explanations, a finding that reinforces other findings in the literature indicating that the relationship between high variability of expertise in small groups and achievement is complex.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This study examines the relationship between "experience grouping" in small groups and learning gains. This chapter recounts the real world problem that led to this line of research, describes the relevance of small group composition to the field of instructional design, and states the research questions.

Overview of the Problem

Johnson and Johnson have characterized cooperative learning as "one of the most widespread and fruitful areas of theory, research, and practice in education…. Cooperative learning is clearly based on theory, validated by research, and operationalized into clear procedures educators can use" (Johnson, Johnson, & Stanne, 2000, introduction). Despite the extensive literature base describing effective cooperative learning techniques, there remain many interesting open questions.

For instance, this research was driven by a small design team attempting to answer a real world problem for which there was no definitive guidance in the cooperative learning literature. Should adults with a wide range of experience who are enrolled in a technical, problem-based training course be grouped heterogeneously, to facilitate tutoring and transfer of experience, or should they be group homogeneously to minimize the possibility that more experienced members would short circuit the learning of less experienced group members by solving the group's problems?

While there is a bias in the cooperative learning literature toward heterogeneous groupings—Watson (1995) noted that "Heterogeneous grouping of students in cooperative learning is so commonly accepted that it is often included as part of the definition of cooperative learning" (p. 292)—in fact, meta-studies indicate that there may be an advantage to homogeneous groupings (Lou et al., 1996). However, there seems to be little empirical research oriented toward adult learners, toward learners engaged cooperatively in solving relatively ill-structured problems, nor toward grouping by experience as opposed to ability as measured by academic assessment instruments. Absent strong guidance, the team decided to engage in original research to find a solution.

Genesis of this Research

In the field of auditing, certified public accountants are brought into a company to state an opinion on the accuracy of the financial statements. Auditors provide reasonable assurance to the readers of those statements that they are free from material error or fraud. Brand new auditors typically take a "new hire" technical training course immediately upon joining their first firm out of college. In these courses, they learn technical and some non-technical skills they will need to become productive members of audit engagement teams, such as how to audit cash accounts, how to interview clients for information, and how to document work performed. Because audit firms make heavy use of summer interns, occasionally an auditor will take the same course twice, once as a new intern and once as a new, permanent audit associate. Susan, an audit associate taking a week-long new hire course for the third time (two internships, then hired) described her experience taking the course the first time:

> I actually cried the first time [I took] it….Because the way that they did it was, it was just like they went over like a whole audit of cash where they were like, "OK, now here's the example, you do it" and I was with three people who had already had internships so they just sat there and were like "OK, we're going to write this here, we're going to do this, and this and this."

Susan was overwhelmed by being the least experienced person in her small group. The group went much faster than she could, leaving her frustrated and deflating her self-confidence. Her group mates were unwilling or unable to provide her the help she needed.

Susan wasn't the only learner that found the small groups in this course ineffective. The evaluation comments submitted by learners included many comments expressing frustration with the small group work, such as:

> I have zero prior auditing experience and therefore didn't find the training as applicable as the others. For the majority of the training I felt like I was in way over my head with the material as the others who have been in the field for some time were pumping out the case studies one after another.

An analysis by the team redesigning this course determined that the number of evaluations specifically mentioning the ineffectiveness of small groups indicated a significant problem with the course.

The design team identified trade-offs in different approaches to experience grouping. On one hand, the team reasoned that heterogeneous groups should be superior because they leverage the knowledge of the more experienced learners—those with practical experience and greater knowledge can act as tutors to less experienced learners. On the other hand, grouping learners into homogeneous groups could allow instructors to concentrate their efforts where they are most needed and would prevent more experienced learners from simply dominating the discussion and doing the work.

Given the evidence of poor group interactivity in the course prior to the redesign, the design team was tempted just to advise instructors to group learners homogeneously . However, the team presumed that group dynamics would vary depending on the course design. The members speculated that increased scaffolding might change the dynamics of heterogeneous groups and help less experienced learners keep up. The design team also intended to implement basic promotive interaction principles from the cooperative learning literature that were lacking in the old course design. For example, the team intended to embed in the course design an explicit emphasis for learners on the expected social functioning of the groups (Johnson & Johnson, 2007).

Previously, the basic course structure had consisted of two largely lecture-based days spent teaching core auditing concepts, followed by two and a half days of large, multiple-hour whole task simulations. There was little in the way of scaffolding bridging the knowledge-based lectures and the skill-based simulations. Around 30 percent of learners who left written comments on the course evaluations asked either for more simulations or for more scaffolding in the existing cases. Comments such as these were typical:

> The trainers would literally tell us cash was next on the agenda, then go sit down. They did not explain any of the [audit program] steps, or go through how to actually perform each step. It’s understandable that as directors they do not do what we as associates do; however, they still could have gone through the

> material better, given a better explanation of what associates do out in the field, and explained the whole purpose of each step. If I had not had training prior to this training, I would have felt extremely lost, frustrated and very turned off by [this firm].

During 2007 and 2008, the course received disappointing results on post-course participant evaluations. The key metric on these evaluations asks participants whether or not they agree with the statement "Overall, I was satisfied with this course," with a one indicating strong disagreement and a five indicating strong agreement. The firm has achieved a historical mean of 4.21 in response to this prompt. In contrast, this course scored only a 4.11 over that same time period, with scores falling from 2007 to 2008. Management at the firm was disappointed that this course should fall consistently below average given its strategic importance as a flagship, high volume course.

This course, like most technical auditing courses at this firm, had been created by internal subject matter experts who had experience delivering instruction but no formal instructional design training. Previous to the redesign effort, the firm hired a full-time instructional designer, the author of this research, who assisted with the course design. The team agreed that, in order to reverse the lack of scaffolding in the old course, basing the new design around the principles of cognitive apprenticeship (Collins, Brown, & Holum, 1991) made sense so the team developed an course architecture built around a "show me, help me, let me try" model. Much more difficult to decide, however, was the issue of group composition. On the one hand, the design team wanted to take advantage of the disparate experience levels present among learners. As it would have been a number of years since any of the instructors had the experience of being a new hire, the team reasoned that there was significant value in having learners hear from each other what to expect on their first audit engagements. Further, the experienced learners would be in a much better position to understand the learning gaps and misconceptions held by inexperienced learners than the teachers would and therefore would be in a good position to fill those gaps and right those misconceptions using language other learners would understand. Finally, since class sizes can be large—over 30 is not unusual—having relatively experienced learners in each group would augment the ability of the instructors

to provide instruction to learners who need it most. The team believed that with the significantly bolstered scaffolding in the course, experienced learners would not have as much of an opportunity to speed ahead of novices and would have more opportunity to recognize their struggles and lend appropriate support. This dual learner/tutor role might even be valuable to advanced learners as having to slow down and explain the content may in and of itself be instructionally impactful.

However, the team worried that the problem of advanced learners racing ahead and leaving the novices behind may be intractable. To place them in the role of both teacher and learner may be unfair to them; or, worse, in a tutor role it is possible they could do actual harm by propagating their own rote understandings or even misconceptions. Having an experienced learner at the table who can quickly solve problems might short circuit rich debate and conversation that might otherwise take place. Further, even though the advanced learners would be in a better position in some ways to understand and address the struggles of novices, there could still be a chasm separating learners with a rich, if incomplete, schema from those who have little schema of how an audit functions.

Further, the team wondered if it was fair to the experienced learners to expect them to come in and be tutors. They debated whether providing them with challenging activities and allowing them to work ahead would allow them to maximize their own growth in the course.

Small Group Composition

The design team consulted various studies in search of guidance. The field of cooperative learning has significant breadth; Johnson and Johnson cataloged the existence of over 900 research studies examining the effectiveness of cooperative learning over competitive and individualistic learning (Johnson, Johnson, & Stanne, 2000, introduction) and themselves have written books aimed at helping classroom teachers implement cooperative learning effectively (e.g., Johnson, Johnson, & Holubec, 1988).

In the view of Johnson and Johnson, cooperative learning represents a key idea in the field: that learners can not only support and teach one another, but that instruction is

more robust, engaging, and effective for this type of interactivity. The principles of cooperative learning have been applied to elementary school mathematics classrooms (e.g., Mulryan, 1995), elementary school students using individualized computer-based instruction cooperatively (e.g., Hooper, Temiyakarn, & Williams, 1993), project-based secondary classrooms (e.g., Cheng, Lam, & Chan, 2008), college psychology classrooms (e.g., Baer, 2003), and asynchronous college courses (e.g., Brewer & Klein, 2006). Johnson and Johnson (2007) cautioned, however, that effective cooperative learning is not a by-product of seating learners together and hoping for the best. Cooperative learning requires taking steps to create a sense of positive interdependence that drives groups to see unflagging cooperation as the key to success for themselves as well as everyone in the group.

Having established the viability and advantages of cooperative learning as an instructional approach, a need for significant refinement in the research base still remains. Mulryan pointed out that "more naturalistic and experimental research is needed to determine the conditions under which these positive outcomes [of cooperative learning] are most likely to occur" (1995, p. 298). With the exception of a 1983 meta-analysis of literature exploring ethnicity- and disability-based diversity in small learning groups (Johnson, Johnson, & Maruyama, 1983), Johnson and Johnson have spent little time in their writings discussing the optimal composition of small groups, preferring instead to focus on cooperative learning methods irrespective of group composition. Though problem solving is at the heart of any small group activity, Johnson and Johnson have also spent little time exploring the types of problem solving best supported by cooperative learning, focusing instead on the generalizability of cooperative learning techniques and positive interdependence principles to new contexts, some beyond the classroom (Johnson, Johnson, & Smith, 2007).

Many researchers in recent years have attempted to isolate various variables that may make cooperative learning more or less effective. Researchers have explored questions around group composition: What differentiates successful groups from unsuccessful groups (e.g., van Boxtel & Roelofs, 2001)? What is the effect of diversity of ability inside small learning groups (e.g., Webb, Welner, & Zuniga, 2001)? What is the

optimal size for small learning groups (e.g., Wheelan, 2009)? What structures and rewards can help infuse cooperative learning groups with positive interdependence (e.g., Slavin, 1996)?

Likewise, researchers have started to monitor the effects of different types of problem solving on questions of cooperative learning (e.g., Merrill & Gilbert, 2008; King, Staffieri, & Adelgais, 1998). Justifying a solution and solution path for an ill-structured problem is different than applying well-structured solution paths to well-defined problems (Jonassen, 1997). What are the effects on group dynamics in cooperative learning of ill-structured problem solving? Traditionally, cooperative learning research as it relates to group diversity has tended to revolve around solving relatively well-structured math and science problems—are the research findings still valid when applied to less well-structured problems (Cohen, 1994)? Are learners with greater expertise better able or less able to assist learners with less expertise under these conditions? The drive to answer these questions is still in its infancy.

One unexplored question in the literature is how well existing guidance in the cooperative learning literature on grouping strategies generalizes to professional, adult learners in real world technical training situations. Are these learners more mature as learners and thus better able to operate effectively in small groups or do they require the same structural techniques as younger learners? Are the motivations of these learners, in courses focused on real world application, where no grade is given, different in ways that affect cooperative learning dynamics?

This study sits at the nexus of these open questions. While this research was designed to answer a specific question about a specific course in a specific industry, the results can be the beginning of a research base for guidance on applying cooperative learning principles to professional, adult learners in training situations.

Specifically, this study will compare the effectiveness of homogeneous and heterogeneous learning groups in a technical training course built around solving a progression of relatively ill-structured problems. For the purposes of this study, a *heterogeneous* learning group is defined as one containing both learners who are experienced in the content domain and those who are inexperienced. *Experienced* in this

study is defined as having been part of one or more audit engagement teams on real world audits.

Effectiveness

Of course, to determine whether one type of group is more effective than another, one first has to specify the variables of interest in making the comparison. The most meaningful metric of effectiveness for this study would ultimately be real world problem-solving performance on the job after completion of the course. Unfortunately, logistical challenges with collecting this kind of data on an appropriate scale to isolate the effects of cooperative learning grouping rendered this path infeasible. Therefore, a proxy was necessary.

Certainly, learning gains are an important part of the effectiveness construct. Learners who gain more applicable skills and knowledge in a course are logically in a better position to be able to solve real world problems upon completion of the course. That said, it is not clear that learning gains as measured by test instruments can fully predict ill-structured problem-solving performance. Jonassen (1997) argued that successful ill-structured problem solving involves creative generation of solutions and cogent argumentation in defense of proposed solutions; it involves reflection. Further, this research will view learning through the lens of the elaboration perspective of cooperation. This perspective posits that cooperative learning is effective because of the nature of the discourse it encourages. For these reasons, for the purposes of this study, *effectiveness* will be defined in part by impact on discourse in cooperative problem solving.

One other important variable is satisfaction. In part, satisfaction is important on a practical level—the firm hosting this study places heavy emphasis on participant satisfaction as a measure of course success, and instructional designers ignore participant satisfaction to their professional peril. Further, if learners overwhelmingly prefer one group type to another, that should weigh in as an important real world consideration. In terms of effectiveness, satisfaction does not guarantee learning—learners may be thrilled with a course that is fun but provides no educational outcomes—but dissatisfaction has been correlated with reducing positive learning outcomes (Winfrey, 1999).

Research Questions

This study will examine the impact heterogeneous and homogenous grouping has on the effectiveness of a course. Specifically, this research will address this question: When professional, adult learners in problem-centered technical training courses are grouped heterogeneously and homogeneously by content experience, what is the effect on course effectiveness as measured by:

1. Learning gains
2. Problem-solving discourse
3. Learner satisfaction

Chapter 2: Review of Literature

The purpose of this study is to examine the impact on learning gains, satisfaction, and problem-solving discourse of grouping professional, adult learners heterogeneously or homogeneously based on domain-level experience. This chapter consists of two major sections. The first section describes the major theoretical perspectives that inform research on cooperative learning. Cooperative learning is a central construct for this research due to the focus on learners solving problems in groups. The second section focuses on research specific to different methods for grouping learners during cooperative learning. Together, the theoretical and empirical background will provide a comprehensive rationale for exploring the questions raised in the previous chapter.

Theoretical Bases of Cooperative Learning Research

Social interdependence theory. Roger and David Johnson (2005) have grounded their research in the construct of social interdependence theory, which they asserted "has had a consistent influence on research and practice since Deutsch introduced it in 1949. It has unusual longevity, scope, and application" (p. 286). *Social interdependence* is a state where the actions of each person in a group affect everyone else in the group in terms of reaching a goal (Johnson, Johnson, Johnson, & Anderson, 1976; Johnson, Johnson, & Maruyama, 1983; Johnson & Johnson, 2005). This construct is distinct from *social dependence*, where one is affected by the actions of the others in the group but the relationship is not reciprocal, and *social independence*, where one is not affected by the actions of others. Johnson and Johnson (2005) also noted a fourth condition, *social helplessness*, where the actions of everyone in the group are completely disconnected from achieving a goal.

Interdependence can be either positive or negative (Johnson, Johnson, Johnson, & Anderson, 1976; Johnson, Johnson, & Stanne, 2000). If the success of everyone in a group depends on everyone in the group being successful, then interdependence is positive. If instead success is zero-sum—an individual can succeed only if one or more other individuals fail—then interdependence is negative. In short, positive interdependence is synonymous with cooperation; negative interdependence is synonymous with competition.

Johnson and Johnson (2005) traced social interdependence theory back to the Gestalt movement in psychology in the early twentieth century. Gestalt researchers framed their research questions around the whole, the *gestalt*, rather than the individual. As Johnson and Johnson (2005) stated, "Gestalt psychologists posited that humans are primarily concerned with developing organized and meaningful views of their world by perceiving events as integrated wholes rather than a summation of parts or properties" (p. 287). Johnson and Johnson next noted the work of Kurt Lewin and the principle of *contemporaneity*, which posits that behavior can only be understood as part of a dynamic social context. As individuals interact, their perceptions change; interdependence in part is a function of individuals perceiving that the members of their group all share motivation to achieve common goals. In the mid-twentieth century, Morton Deutsch built on Lewin's theories to differentiate positive and negative interdependence (Johnson, Johnson, & Smith, 2007). Deutsch saw the distinction between positive and negative interdependence as a continuum. A productive action in positive interdependence situation benefits everyone. In a negative interdependence situation, individuals are in competition; an action that is productive for one individual is necessarily a setback for another individual.

According to social interdependence theory, learners have goals and take actions to achieve them. If learners are positively dependent on others, they tend to take promotive action. If learners are negatively dependent on others, they tend to take oppositional action. The more promotive action, the greater overall achievement. Cooperative learning, therefore, should lead to more powerful learning than competitive learning (Johnson, Johnson, & Smith, 2007).

Even though social interdependence theory predicts a superiority of cooperative learning over competitive, the theory also predicts a few scenarios where cooperative learning can break down—Johnson and Johnson (2005) referred to these as "pathologies of cooperation." One principle of social interdependence theory is *substitutability*, which posits that an advantage of positive interdependence is that is creates efficiencies—work does not have to be duplicated the way it does in competitive situations. However, substitutability can have a negative effect when it leads to individual group members

becoming too specialized. Another principle of the theory is *cathexis*—the idea that individuals in cooperative groups will develop social bonds that will further encourage positive achievement. Cathexis can lead to cliques, though, which can undermine group effectiveness. *Inducibility*, the principle that cooperative learning creates an atmosphere of trust and influence, can also theoretically backfire by leading to groupthink.

Johnson and Johnson (2005) acknowledged several limitations of social interdependence theory. First, the theory treats human actors and groups of actors as being single-minded in the pursuit of goals. At any given moment, humans act upon an array of conscious and subconscious complimentary and contradictory goals, the complexity of which is difficult to capture in a model. Further, whereas the theory tends to assume individuals act in their own self-interests, that is not always the case. Also, goals are driven in part by history—individuals who have worked with each other before will naturally be affected by past interactions. A side effect of this limitation is that most empirical research on social interdependence theory centers around subjects who have never before met (a condition that is generally true in the present study as it centers on a new hire course). Second, Johnson and Johnson pointed out it is not clear how well social interdependence theory scales up to large group sizes. Third, the theory says little about the power differentials that exist in any group and the effect they have on the actions of group members.

Social-behavioral perspectives. The Johnsons' model of cooperative learning has been characterized as a *social-behavioral perspective* (Slavin, 1996; O'Donnell & O'Kelly, 1994). Slavin (1996) identified two major approaches to designing cooperative learning within the social-behavioral perspective: the motivational approach and the social cohesion approach.

Table 1. Major Approaches to Designing Cooperative Learning Activities

	Social-behavioral		Cognitive	
	Motivation	**Social cohesion**	**Elaboration**	**Developmental**
Focus	Structural rewards drive cooperative behavior	Sense of team creates intrinsic incentives to work cooperatively	The act of talking encourages cognitive restructuring and refinement	Leveraging natural learning relationships between peers and/or mentors
Associated researchers	Slavin	Johnson & Johnson	Pressley Teasley Roschelle	Piaget Vygotsky

The motivational approach. The motivational approach to cooperative learning is associated with the work of Robert Slavin (O'Donnell & O'Kelly, 1994) and is exemplified in STAD: Student Teams-Achievement Divisions. STAD is a form of cooperative learning where all learners in a cooperative learning group are evaluated separately, but each person's grade depends on the performance of everyone in the group (Slavin, 1996). If one person in a three person group does well on an assessment but the other two do poorly, no one gets a good score. This model contrasts with the cooperative models often employed in classrooms today, where either groups create one project that determines the grade of everyone in the group, and which encourages groups to rely heavily on the most advanced learner, or groups work together but are evaluated separately—which also provides little external incentive for more advanced learners to take charge of helping struggling learners master the material. The motivational approach posits that learners will become motivated to help each other if their own performance will be in part evaluated by everyone else's individual learning gains, and that this increased help will lead to greater overall learning gains for the group.

Slavin contended (1996) that the motivational approach is only necessary when intrinsic motivation to ensure that everyone in a learning group makes significant learning gains is low, which may often be the case in typical classroom settings. Logically, then, it may not be appropriate in other situations, such as voluntary learning

groups, where intrinsic motivation may be higher. There is also some evidence that providing external motivation along the lines of STAD may be less necessary for higher level problem-solving objectives than for lower level objectives (Cohen, 1994).

Social cohesion approach. The other approach associated with the social-behavioral perspective is the *social cohesion* approach (Slavin, 1996; O'Donnell & O'Kelly, 1994). This approach to designing cooperative learning concentrates on helping cooperative groups achieve a sense of cohesion so that external reward systems are not necessary. This approach is more likely to appeal to those uncomfortable with potential side effects of extrinsic motivation (e.g., Kohn, 2004). The focus of the social cohesion approach is on creating intrinsic motivation, on fostering a feeling of camaraderie and togetherness that makes group members feel invested in the success of the other members. Social cohesionists are likely to engage cooperative learning groups in team-building activities or group reflection activities. Jigsaw and related activities, which create interdependence by giving everyone in the group unique information needed by other members, are associated with this perspective. Slavin (1996) was critical of pure social cohesion approaches, arguing, "In general, methods which emphasize teambuilding and group process but do not provide specific group rewards based on the learning of all group members are no more effective than traditional instruction in increasing achievement" (p. 48). However, this assertion may be at odds with the Johnson, Johnson, and Stanne meta-analysis (2000) that shows significant positive effect sizes for a variety of cooperative learning techniques that do not necessarily involve external reward structures.

Social-behavioral perspectives on cooperative learning in adults. The social-behavioral perspective on cooperative learning is an area for investigation that appears to have largely ignored post-college adult learners. In 2007, Johnson, Johnson, and Smith authored a review of social interdependence theory entitled "The State of Cooperative Learning in Post-Secondary and Professional Settings." The article talked in some detail about cooperative learning at the college level but said very little about post-college professional training; it appears that by "professional settings" the researchers were referring mainly to post-secondary professional schools.

In the article, Johnson, Johnson, and Smith (2007) drew no distinctions between how social interdependence theory applies to adults and how it applies to children; that is, they did not discuss how maturity affects social interdependence, leaving open questions of whether adults can utilize their maturity to achieve social interdependence more easily. The article is similar to other summary articles of social interdependence theory Johnson and Johnson have written in recent years with the exception that the supporting evidence is culled from studies focused on subjects 18-years-old and older. For example, when the researchers asserted that cooperative learning leads to greater effort to learn than competitive or individualistic learning, their analysis was very similar to a 2005 Johnson and Johnson article about current developments, not specific to a particular age range, related to social interdependence theory. They described the number of studies (smaller for adults, but still large at 150) and the effect sizes (around 0.65 for all ages, around 0.50 for adults), but there was no analysis in either case around the role of maturity in cooperative effort. The implication is that the psychological processes underlying social interdependence are the same no matter the target population, or at least that the differences do not matter, though this is never expressed explicitly.

While Johnson and Johnson have not attempted to tease out the effects of experience and maturity on social interdependence, they did describe in their 2005 general overview some ways that social interdependence theory has been or can be applied to business and industry settings. These applications center around work teams and conflict resolution, however, and not particularly around training.

In assigning a central role to social forces in trying to understand better adult motivations as they relate to the workplace, social interdependence theorists are among researchers characterizing workers as being driven by highly complex, contradictory, social impulses (Murtonen, 2008) that mix intrinsic and extrinsic motivations (Alejandro, 2002). As Murtonen (2008) stated, when it comes to analyzing the professional motivations of adults, “it is important to be cautious and avoid overly simple conclusions about causality” (p. 220-1). There are efforts underway to understand better the motivations of professionals as they relate to professional development. Karen Jensen (2007), for instance, has studied how accountants, among other professionals, view the

demands of professional development as their field undergoes increasingly rapid change and innovation. Jensen defended the utility of using a social, motivational lens to study this problem as more cognitive approaches "overlook the role that excitement, passion and desire may play in the learning process. They focus on the how and what of learning and overlook the 'why' questions that drive the learning process" (p. 490).

The cognitive perspective on cooperative learning. In contrast to the social-behavioral perspective stands the cognitive perspective (O'Donnell & O'Kelly, 1994; Slavin, 1996). As O'Donnell and O'Kelly noted, "Social psychological approaches to cooperative learning emphasize social mechanisms that promote interdependence. In contrast, cognitive approaches to cooperative learning emphasize mechanisms that promote cognitive interdependence for cooperating group members" (p. 333). O'Donnell and O'Kelly outlined two major strands in the cognitive perspective: the developmental perspective and the elaboration perspective.

The developmental perspective. O'Donnell and O'Kelly (1997) noted that developmental perspective has been framed by contrasting Piagetian and Vygotskian views on learning. Piaget argued that all significant intellectual advances proceed through a sequence of equilibrium, disequilibrium, and re-equilibrium (Flavell, 1996). Disequilibrium is the condition created when learners are introduced to a puzzling or contradictory aspect of their existing knowledge, forcing some change in understanding in order to establish re-equilibrium. According to Piaget, who was specifically interested in the learning of scientific principles, the purpose of this cycle is to help individuals narrow the gap between their own flawed understandings and correct scientific reasoning—a position much commented on by critics who see this hierarchy of knowledge as universalistic rather than relativistic (Matusov & Hayes, 2000).

Disequilibrium can result from any number of catalysts, such as an observation that does not match one's mental model. In cooperative learning situations, interacting with a peer who has a different understanding could be a source of disequilibrium (Webb, 1991; Dorval, Eckerman, & Ervin-Tripp, 1984). Piaget saw the cooperation of equal partners as an important source of new perspectives and a catalyst for reflection (Matusov & Hayes, 2000). While cooperating, learners would discover contradictory perspectives,

creating disequilibrium, and work together to resolve the contradictions and establish re-equilibrium.

Vygotsky, in contrast, focused on how individuals acquire new knowledge by interacting with individuals who know more than they do about the subject. He argued that more knowledgeable mentors can help individuals bridge their zones of proximal development (Webb et al., 2008).

In significant ways, Piaget and Vygotsky were at odds in their views of the role of sociocultural forces in learning (Bockarie, 2002). Bridging Vygotsky's zone of proximal development (ZPD) requires the presence of someone with more advanced knowledge to provide the scaffolding. Piaget, on the other hand, saw significant danger in the presence of the more advanced individual (Matusov & Hayes, 2000). Learning in his view requires experience and reflection, and the presence of a more experienced mentor is likely to abrogate these necessary conditions. Piaget called this phenomenon *social constraint*. In his view, the more advanced individuals will impose their knowledge on the less advanced learners in a way that might produce a shallow, temporary understanding, but it will fail to induce true transformation. When it comes to cooperative learning, then, Piaget would seem to argue in favor of homogeneous grouping, while Vygotsky seems to be arguing more in favor of heterogeneous grouping.

Given those basic positions, it would be easy to portray a bright line between Piaget and Vygotsky—indeed, they are often portrayed in opposition in small group learning literature (e.g., Webb & Mastergeorge, 2003)—but there were many areas of overlap and nuance (Matusov & Hayes, 2000). First of all, Piaget's views around cooperative learning were complex. He did not believe that a group of peers in a learning situation automatically became a high functioning cooperative group. To achieve the kind of dialog necessary for disequilibrium and re-equilibrium, he believed learning groups required diversity of perspective, critical dialog, and mutual reflection. In the same way, he did not believe that heterogeneous learning groups automatically led to social constraint; he further acknowledged that inequality in learning groups can be a positive development experience for the more advanced learner.

Further, it would be an oversimplification to say that Vygotsky rigidly believed that a learner requires a mentor of greater knowledge in order to scale the zone of proximal development. Vygotsky believed not only that interaction with equals was a path to developing with a ZPD, but even that an activity without peer interaction can help learners grow (Matusov & Hayes, 2000).

From the perspective of researchers interested in adults learning cooperatively, a weakness of the developmental perspective is its focus on young learners. Piaget, for example, was interested in studying children in part because he was driven to explain the apparent differences he saw between how children think and how adults think (Ackermann, 2001), making the application of his ideas to adult learners problematic. Vygotsky, too, derived his theories on bridging the ZPD from his work on child development—he even framed the ZPD in terms of children, defining it as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1978, as quoted in Reynolds & Fletcher-Janzen, 2009).

Ideas based in developmental theories have, however, been applied directly to adult learners. Bockarie (2002), for instance, argued that Vygotsky envisioned learning as a lifelong process and therefore his ideas, particularly as they relate to the ZPD, can be reasonably applied to adult learning processes. Bockarie applied them in detail to informal, vocational adult learning processes though he did not attempt to tease apart the implications of maturity, nor did he articulate a theoretical position that bridges a developmental theory to adult processes. A longstanding, unresolved debate in psychological circles concerns whether *development* and *learning* are distinct processes or two names for the same process (Packer & Goicoechea, 2000). If they are indeed psychologically the same, that could arguably open doors for applying developmental theories to adult learning situations.

Bandura's social cognitive theory might be one path to applying the core principles of developmental theories to adult learners while at the same time addressing the apparent dichotomy between the Piagetian and Vygotskian positions. Social cognitive

theory posits that human beings are agents capable of influencing their own circumstances and responses (Bandura, 2004, 2006). In other words, people self-organize and self-regulate. Social cognitive theory is based around four principles. First, that people act intentionally. Second, that people engage in forethought and planning. Third, people are self-reactive, which means that regulate themselves in the face of feedback from the environment. Fourth, people are self-reflective. They look back and use what they learn to make further plans. Despite these principles of autonomy, social cognitive theory emphasizes that people are never wholly autonomous agents. Their actions are socially situated. From the perspective of studying adult learners, a significant advantage of social cognitive theory is its post-developmental focus (Martin, 2004).

One of the central constructs associated with social cognitive theory is *self efficacy*. Self efficacy theory describes how the performances of individuals are affected by their own beliefs in their ability to succeed at a task by themselves. *Collective efficacy* in contrast describes how the success of a group is influenced by its collective belief in its abilities (Bandura, 2000, 2006; Lent, Schmidt, & Schmidt, 2006)—although the concept itself has been criticized as being inconsistently defined (Maddux, 1999; Paskevich, Brawley, Dorsch, & Widmeyer, 1999). Bandura (2000) called collective efficacy a "group-level property," describing the ability of the entire group to achieve a goal or solve a problem. As Bandura pointed out, the collective efficacy of the Chicago Bulls was presumably a lot higher when Michael Jordan was in the lineup than when he was absent. A group of low-skilled individuals with one superstar may have high collective efficacy about their ability to solve a given problem even though the superstar may do the bulk of the problem solving and, perhaps, therefore, the learning. As a group-level property, this construct is therefore not designed to describe or predict the performance of any individual member of a group.

Cheng, Lam, and Chan (2008) had a different take on the issue. They argued that group composition can be reflected by measures of collective efficacy. For example, if a high performer is in a small group with low performers, the high performer might feel that he or she is disadvantaged by the low performers and collective efficacy suffers. Likewise, homogeneous groups of low performers are also likely to have low collective

efficacy, and low collective efficacy correlates with motivation, commitment, and achievement (Bandura, 2000). Or, in contrast, low performers grouped with high performers might feel more confident, resulting in higher collective efficacy, leading to higher group cohesion and greater accomplishment for all the learners. Cheng, Lam, and Chan referenced Vygotsky's views as supportive of this possibility. Still, a significant challenge in applying collective efficacy in particular to small group learning is the potential for confounding the achievement of the group goal with individual learning. A group's ability to achieve a common goal does not necessarily correlate with learning across all learners in the group.

Bandura's theories may help lay a theoretical foundation for exploring cooperative learning by potentially resolving the perceived Piaget/Vygotsky dichotomy. Bandura's social cognitive theory has been called "a possible alternative to Vygotskian socioculturalism and Piagetian constructivism, one that might help to explain how agents can be both socially constructed and agentically constructive" (Martin, 2004, p. 140). In other words, if Piaget is associated with individual experience being the key to learning, and Vygotsky with social scaffolding being the key, Bandura provides a middle position by suggesting in social cognitive theory that individuals are agents that both inform and are informed by a social context that is itself made up of agents. Learners simultaneously act individually while being influenced socially.

Packer and Goicoechea (2000) argued that the differences between the socio-cultural perspective associated with Vygotsky and the constructivist perspective associated with Piaget are at their heart ontological. Ontologically, constructivism posits that the real world exists but can only be individually inferred. Socioculturalism, however, is predicated on the assumption that there is not a strict demarcation between the self and the world, and that knowledge therefore is a complex, distributed social phenomenon.

Efforts to marry socioculturalism and constructivism have been made beyond Bandura. Packer and Goicoechea (2000) argued that a middle position, a "nondualist ontology" can reconcile the two perspectives into complementary theoretical positions where learning represents both personal and social transformation. They went on to

describe six themes of a nondualist ontology: “the person is constructed, in a social context, formed through practical activity, and in relationships of desire and recognition that can split the person, motivating the search for identity” (p. 239). People are formed socially but at the same time they make choices that in turn transform society. To be a person is to be divided between the individual and the social. We seek to belong but we also seek to establish a unique identity.

The elaboration perspective. The second major cognitive perspective on cooperative learning is the *elaboration* perspective (O’Donnell & O’Kelly, 1994; Slavin, 1996). From this perspective, the positive effects of cooperative learning are highly dependent on the act of elaborating a position. O’Donnell and O’Kelly (1994) described it this way: “As a result of interaction, peers elaborate their cognitive structures. For example, a peer may provide an example for a concept that helps elucidate that concept. ‘Elaboration’ can involve the restructuring of existing cognitive structures or the addition of new information to existing structures….The role of verbalization in the cognitive/elaborative approach is critical” (p. 338).

The role of talk is an area of commonality with Piaget and Vygotsky, specifically in their interest in the relationship between thought and language (Carter & Gail, 1994). Both researchers were interested in a phenomenon known as *private speech* (Teasley, 1995), which refers to the spontaneous act of talking to oneself while engaged in an activity. In her review of the literature in role of private speech in problem-solving success, Teasley (1995) pointed to a positive long- and short-term benefit to private speech while problem solving, specifically private speech that consists of interpretations of strategies, concepts, and principles. Teasley used this framework as a base for her research in cooperative learning by hypothesizing that it is this articulation of interpretation that makes cooperative learning a powerful learning strategy—cooperative learning encourages talk and explanation and therefore should lead to greater learning.

Such a position fits well with the view that learning is deepest when it is generative, that when learners have to elaborate on materials, they end up understanding and remembering them better (King, 1992; Wood, Pressley, & Winne, 1990; Pressley, Symons, McDaniel, Snyder, & Turnure, 1988). Spontaneous elaboration by learners is

relatively rare when they are alone (King, 1992; Pressley, et al., 1992), which again may suggest why cooperative learning fares so well in studies since it encourages interaction.

If talking correlates with learning, it follows that those group members who talk the most should learn the most (Cohen & Lotan, 1995). Even within a single small group, some learners have greater opportunity or encouragement to talk than others. Expectation states theory predicts that every member of a group has a particular status, and that status affects how individuals act and are acted upon in the social context. Though status can shift (Fişek, Berger, & Norman, 1991), especially as a group initially becomes established, it profoundly affects interaction patterns, with high status individuals enjoying exclusive privileges. When a small group is formed, group members immediately begin to form performance expectations, often unconsciously, for each other based upon initial performance as well as markers like age and race (Kalkhoff & Thye, 2006). It is these perceptions that drive some of the group dynamics in terms of whose opinions the group seeks out, who gets the most encouragement to speak, and so forth (Cohen & Lotan, 1995).

While the amount of talk offered by an individual in a cooperative learning situation may be important, the complete picture is almost certainly more complex—after all, cooperative learning is not just a group of individuals engaging in parallel private speech. As noted above, Vygotsky among others believed that the process of social construction of knowledge is complex. At least one cooperative learning researcher (Roschelle, 1992) has based his research on a relational theory of meaning that has its roots in social constructivism. This approach posits that social interaction enables learners to construct meaning incrementally. Learners gradually refine their understanding of concepts—in Roschelle's case, science concepts—through social interactions. In this theoretical view of learning, conceptual understanding undergoes a convergence among learners through an iterative process of displaying, confirming, and repairing shared meanings. Thus, it may be that while *quantity* of speech is important, *quality* of speech is also likely to be important. Speech or other interactions that encourage this iterative convergence are likely to enhance learning. Logically, interactions that inhibit these processes can be counterproductive.

Problem-solving structure: A fifth approach. Slavin (1996) noted that not all four perspectives described above are appropriate for all cooperative learning situations—he suggested, for example, that cognitive perspectives may be more appropriate for pairs of learners rather than larger groups, as "pairs involve a much simpler social process than groups of four" (p. 52), though the principles of social cognitive theory suggest that even the interaction of pairs can be highly complex (Bandura, 2004, 2006). In any case, Slavin saw the four perspectives as complementary rather than contradictory.

One commonality among the four approaches is that they all focus on the interactions of group members in trying to understand the conditions that make cooperative learning effective, regardless of the type of problem the group is trying to solve. The four approaches, however, offer little in the way of theoretical guidance as to how different types of problem solving promote or discourage interactivity. Does higher-level problem solving—learning activities, for example, at the higher levels of Bloom's taxonomy (Anderson, et al., 2000)—promote different group dynamics than activities at the lower levels of the taxonomy? Does it matter how higher level activities are structured?

David Merrill and Clark Gilbert (2008) contended that in fact peer interactivity is heightened when peer groups are challenged to solve complex real world problems that are presented in an instructionally meaningful progression. Here they drew a distinction between *problem-centered* and *problem-based* instruction. The latter, they contended, is synonymous with giving groups of learners complex problems along with necessary resources but little scaffolding. Problem-centered instruction, in contrast, offers a carefully sequenced progression of problems designed to support learners systematically. Included among the advocates of problem-centered instruction is van Merriënboer and his 4C/ID model of instructional design (van Merriënboer, 1997; van Merriënboer & Kirchner, 2007).

Problem-*based* instruction, Merrill and Gilbert (2008) argued, is typically topic-based: learners receive instruction in a number of topics important for the problem-solving task and then the whole skill problem-solving practice is the cumulative activity

where learners are asked to put it all together. They criticized this approach because, not having had the problem-solving experience yet, learners are not able to map the information taught prior to problem solving to meaningful mental models that will effectively help them in the problem-solving task. Instead, learners are forced to rely on associate memory to record and recall the decontextualized skills and knowledge. Problem-centered instruction overcomes this limitation by couching the instruction inside of problem-solving iterations. A useful mental model of the problem-solving task is established from the beginning and is continually refined and elaborated. New skills and knowledge can be mapped directly to the appropriate mental model rather than stored in associative memory.

While nothing in the problem-centered instruction model mandates cooperative learning, Merrill and Gilbert (2008) asserted that interaction with peers helps learners tune and stabilize their mental models by forcing learners to process information more deeply and to examine their assumptions more closely in order to be able to articulate their understandings to their peers.

To that extent, Merrill and Gilbert's position can be seen as a means to operationalize the elaboration perspective on cooperative learning. If the elaboration perspective contends that cooperative learning produces optimal results when learners spend significant time articulating their mental models, then Merrill and Gilbert's contention dovetails by providing an instructional sequence designed to heighten and deepen that dialog. Further, Merrill and Gilbert contended that problem-centered instruction inherently supports cooperative learning. Problem-centered instruction, they noted, provides "needed structure and guidance for effective peer interaction" (p. 203) that, where absent, as in problem-based instruction, leads toward what they call *peer-telling*, or the one-way dissemination of information from one peer to another as opposed to real dialog. They continued, "Peer interactions are also facilitated when there is a carefully designed progression of problems because tuning each new problem forces learners to accommodate their mental model to the new situation and the peer-collaboration or critique causes them to tune their mental model to be more flexible to

adapt to the variations that may exists for doing each task or solving each problem in the sequence" (p. 203).

In their conception of problem-centered instruction, Merrill and Gilbert leaned heavily on the work done by van Merriënboer, Kirchner, Sweller, and others who advocate a scaffolded, iterative problem-solving instructional model for learning to solve complex real world problems. According to van Merriënboer and Kirchner (2007), to be effective such an environment has to have four interrelated components: learning tasks, part-task practice, supportive information, and procedural information. *Learning tasks* are authentic, whole-skill problem-solving experiences that are organized into task classes and sequenced purposefully. Learning tasks are supported by scaffolding and consist of recurrent and non-recurrent skills. Recurrent skills—those that are used over and over again and require a level of automaticity in order to free cognitive load and add efficiency to problem solving—are practiced in *part-task practice* in order to build that automaticity. Part-task practice is the only part of the model where skills are used out of context. Learning tasks are supported by two types of resources: *supportive information* and *procedural information*. The former provides strategic guidance to learners in solving the problems, the latter provides tactical information about how to complete well-defined procedures.

While Merrill and Gilbert (2008) suggested that careful sequencing of problems promotes dialog inherently, they also suggested ways to augment the basic design specifically for cooperative learning. They first suggested encouraging learners to share relevant prior experiences with each other as part of the learning sequence. These activities fulfill Merrill's *activation principle*, which states that learning is enhanced when relevant cognitive structures are first activated by recalling prior knowledge or experiences (Merrill, 2002). While the activation principle might suggest that this kind of exchange is most likely to benefit more experienced learners in a heterogeneous group, Merrill and Gilbert (2008) assert that these exchanges should help listeners as well by giving them "vicarious experience which in turn activates their previous experience and associated mental models" (p. 204).

Merrill's *demonstration principle* states that learning is promoted when learners can observe the skill in action (Merrill, 2002). This can take the form of written out worked examples. Merrill and Gilbert (2008) suggested that supporting examples with challenging, thought-provoking questions and then asking learners to discuss them is an effective collaborative technique because taking part in those discussions requires learners to tune their mental models. Merrill's first principles also state that effective instruction encourages learners to apply new knowledge to real world problems; this is the *application principle*. Merrill and Gilbert asserted that asking learners to make this application collaboratively is effective, though on this point they do not articulate how collaborative learning is superior to individual learning. Finally, Merrill's *integration principle*, which states that "learning is promoted when learners integrate their new knowledge into their everyday life by being directed to reflect on, discuss, or defend their new knowledge or skill" (Merrill and Gilbert, 2008, p. 204), can be applied in collaborative situations when learners engage in peer critique.

Merrill and Gilbert's article is provocative because they approached cooperative learning from a different angle than most researchers. In keeping with Slavin's framework described earlier, most cooperative learning research focuses on structuring cooperation in ways that are not specific to any particular content or instructional approach. For example, Slavin's STAD is generic in the sense that it is not meant to be applied to any particular subject area or level of Bloom's taxonomy. Merrill and Gilbert, however, approached the problem from the other side, focusing less on the structure of cooperation and more on the learning activities themselves, suggesting that well-designed problem-centered instruction lends itself naturally to effective cooperative learning—that effective cooperation is a by-product of problem-centered instruction, especially with a few additions such as encouraging peer critique. This is not to suggest that Merrill and Gilbert are asserting that specific cooperative learning techniques are not necessary or useful, but rather it suggests that when it comes to cooperation, the place to start is with the design of the instruction and not the design of the cooperation itself. In other words, most cooperative learning research focuses on how to graft cooperative learning onto existing learning activities; Merrill and Gilbert instead appear to encouraging teachers

and instructional designers to create learning experiences that intrinsically encourage cooperation.

Merrill and Gilbert's article may be viewed as a nudge to the research community to validate their assertions. While a great deal of research has gone into problem-centered instruction (van Merriënboer & Kirchner, 2007; van Merriënboer, 1997; Kirschner, Sweller, & Clark, 2006), Merrill and Gilbert's extension of problem-centered instruction into cooperative learning is largely at the assertion level. They present no references to empirical evidence that their ideas extrapolate to cooperative learning situations, instead they rely chiefly on the core problem-solving literature, Merrill's first principles of instruction, and van Merriënboer's 4C/ID, none of which are specific to cooperative learning.

Other research into cooperative learning and problem solving. While much of the cooperative learning research base focuses on structures of cooperation rather than on structures of problem solving, Merrill and Gilbert were not the first to explore the relationship between problem-solving activities and cooperative learning. Song and Grabowski (2006) framed the issue as one of goal orientation. From this perspective, environments that encourage learners to adopt a learning orientation generally produce superior motivation and learning than environments that instead encourage them to adopt a performance orientation, especially in ill-structured learning environments. While in their research with middle school students they found mixed results, they recommended that in cooperative learning environments focused on ill-structured problem solving that instructors include elements that encourage learners to adopt a learning orientation, such as support for significant learner control and inclusion of task-instruction messages that remind learners to focus on understanding rather than achievement. They further recommended avoiding elements that encourage performance orientations, such as placing teams in competition. To that extent, Song and Grabowski were analyzing cooperative learning from a social cohesion perspective.

Kumpulainen and Kaartinen (2003) used a qualitative, "microlevel" analysis of collaborative reasoning to better understand how open learning tasks affect cooperative dialog. They noted that their research was driven by the assumption that closed and rote

learning tasks produce fundamentally different collaborative dynamics and that very little is known about the microlevel processes of peer collaboration in ill-structured problem-solving tasks. This level of analysis is important in part because, as noted above, while quantity of speech may be important in cooperative learning, quality is also important. Kumpulainen and Kaartinen noted that not all talk enhances cooperative learning—talk can degrade learning when learners propagate ideas that are detrimental to solving the problem. They analyzed three cases of collaborative problem solving by elementary school dyads working on a geometry design problem and painted a highly complex picture of problem solving. "The strategies used in problem solving included versatile approaches consisting of constructing, measuring, visualizing, testing, excluding, and evaluating. The students' verbal interactions included reasoning, hypothesizing, arguing, organizing, and questioning" (p. 366). However, when there was not a "shared focus and explicit attempts to maintain collaboration with in the peer group" (p. 366), then the dyads often suffered from conflict and asymmetric interaction. Acknowledging their work as merely a first step, they concluded, "This study shows that the social and cognitive processes inherent in peer problem solving are highly dynamic in nature, shaping collaborative reasoning on a moment-by-moment basis" (p. 367). One potential critique of their research from the perspective of Merrill and Gilbert's work is that the problem-solving design task, creating three dimensional objects from two dimensional pictorial representations, may not represent a real world, relevant problem-solving task to the target population.

Kumpulainen and Kaartinen's work in turn is based in part on the work of Elizabeth Cohen. In her 1994 conceptual review of the literature related to the conditions under which small group learning is productive, Cohen suggested that the distinction between problem solving and more routine learning is an important one for the field: "Whereas limited exchange of information and explanation are adequate for routine learning in collaborative seatwork, more open exchange and elaborated discussion are necessary for conceptual learning with group tasks and ill-structured problems" (p. 1). Further, she drew a distinction between group tasks—those activities that require "resources (information, knowledge, heuristic problem-solving strategies, materials, and

skills) that no single individual is likely to solve the problem or accomplish the task objectives without at least some input from the others" (p. 8)—with more routine learning tasks where the learners are working cooperatively on tasks that they would normally complete alone. Grafting cooperative learning onto traditional learning tasks tends to be more about allowing stronger students to help weaker students, creating a sort of unilateral interdependence. Weaker students depend on the stronger students but not vice versa. Complex instruction, she noted, anticipating Merrill and Gilbert, creates conditions of reciprocal interdependence. She expresses this as a "general proposition: Given an ill-structured problem and a group task, productivity will depend on interaction" (p. 8).

Cohen (1994) was critical of Slavin's emphasis on external motivation. She contended that his perspective

> would appear to apply better to the kinds of collective or collaborative seatwork tasks that are so common in cooperative learning, where it is of vital importance to motivate those would could do the task by themselves to interact and assist those who are having difficulty. These are not group tasks, as defined earlier in this article, because they could be carried out by one individual. Reward interdependence does not appear to be necessary for achievement when students are motivated to complete a challenging and interesting group task that requires everyone's contribution for a good outcome. (p. 15)

She concluded:

> This analysis moves away from the fruitless debate about intrinsic and extrinsic rewards and goal and resource interdependence that have tied the field into theoretical and ideological knots for some time. Instead, I have recommended that propositions be conditionalized on whether or not the assignment given to the group is a true group task and whether or not it is a problem with an ill-structured solution. (p. 30)

Cohen, and more recently Merrill and Gilbert, have been reframing cooperative learning from a set of techniques to be retrofitted onto existing, traditional instruction to a foundational instructional design approach to teaching ill-structured problem solving.

Structuring interaction. Johnson and Johnson (2007) made clear their contention that simply putting learners together in groups and allowing them to self-organize or self-regulate, or not, does not constitute cooperative learning. For formal learning to be meaningfully cooperative, they argued, appropriate elements must be structured into the situation.

The social-behavioral approach to cooperative learning is focused in large part on imposing structured activities that foment either motivation (e.g., STAD) or group cohesion (e.g., teaching groups conflict resolution strategies). Researchers looking at cooperative learning from the cognitive perspective tend to focus more on structuring the interactions in ways that promote optimal cognition, such as by assigning learners individual roles in order to promote the exploration of different perspectives.

The insights of Cohen and others as to the role of problem solving in cooperative learning, however, surfaces the question of how much and what types of structure are necessary when the learning event is a carefully sequenced progression of ill-structured problem solving. Cohen (1994) observed, "Herein lies the dilemma: If teachers do nothing to structure the level of interaction, they may well find that students stuck to a mostly concrete mode of interaction. If they do too much to structure the interaction, they may prevent the students from thinking for themselves and thus gaining the benefits of the interaction" (p. 22). It is a balancing act, and there is no coherent picture in the literature of how much structure is appropriate under what conditions.

Types of structures to support appropriate levels of interaction include setting ground rules for help-seeking (e.g., Nattiv, 1994) or even providing cue cards or think sheets to remind groups to use behaviors that drive toward deeper discussions (e.g., Mercer, Wegerif, & Dawes, 1999; Meloth & Deering, 1994). Learners can be trained in asking and answering more open-ended questions (King, 1992). Learners can be taught to assume specific roles, such as in reciprocal teaching (Palincsar & Brown, 1984), which asks learners to take turns assuming the role of leading a dialog.

While structuring the level of interaction has had many successes in the literature, there has been no systematic analysis along a couple of important dimensions. One unexplored dimension is the function of maturity. Published studies that impose discourse

structures on learning groups are heavily skewed towards younger learners. This emphasis may represent a tacit assumption that the importance of these kinds of structures fade as learners become more mature.

A related issue here is the level of effort necessary to train teachers to use structured learning techniques. Nattiv (1994), for instance, had teachers go through 18 hours of instruction on cooperative learning techniques before beginning her research. Webb, Troper, and Fall (1995) had participating teachers complete several full day workshops before implementing cooperative learning for their study. Gillies and Khan (2009) had teachers complete two days of professional development. In technical training in the audit field, where the instructors are often subject matter experts who only teach very occasionally, putting in place structures that require extended training may not be feasible.

The other unexplored dimension is whether structure is more or less needed based on the heterogeneity of the group. Saleh, Lazonder, and de Jong (2007) suggested that structure, at least in the form of detailed scripts, "is largely removed when students are of different ability or achievement level" (p. 316) as under these conditions the more advanced learner naturally assumes a tutor role and begins engaging in the types of activities often provided by the external structure: providing explanations, giving feedback, prompting for action, and so forth. The analogy to peer tutoring may be problematic, however, as the dynamics of dyads can be much different than groups of three or more (Lou et al., 1996). Still, Saleh and colleagues raise a point: it would be reasonable to expect the discourse in a heterogeneous group to be different, all else being equal, than the discourse in a homogeneous group. The differences may warrant different kinds of instructional structures to support optimal learning.

Empirical Research into Group Composition Effects

Group composition. In 1996, Lou et al. published the largest meta-analysis to date of the small group learning research base, covering more than 500 studies meeting their selection criteria. They concluded that, on average, students learn more in small groups than they do in whole class activities, especially when the small groups utilize cooperative learning techniques. However, the effect sizes in the literature show high

variability; while many studies reported significant effect sizes above zero, many also reported effect sizes of zero or less, suggesting that further analysis is warranted to isolate the variables causing the heterogeneity in the results.

The researchers' post hoc analysis isolated a number of significant factors. The largest source of variability was in how outcomes were measured; studies utilizing researcher- or, especially, teacher-prepared tests were much more likely to show results in favor of cooperative learning than results from standardized tests. This source of variability is difficult to interpret because it can indicate that those tests were more sensitive to the learning objectives targeted in the studies, but it can also indicate systematic bias in favor the of experimental condition. Other significant factors included:

- Effect sizes were significantly higher when instruction and materials were modified for the different conditions. In other words, when comparing small group versus whole class learning, one would expect that the small groups would do better if the materials were adapted specifically for small groups rather than both conditions using the exact same materials.
- The more training given to teachers in the cooperative learning technique, the greater the learners' achievement.
- Smaller groups (three to four) achieved more than larger groups. By the time group size grew to about six or more, achievement became equivalent with traditional whole class instruction. This conclusion parallels similar findings for the productivity of teams in the workplace (Wheelan, 2009).
- The effects of small group instruction were greater for math and science classes than language arts and other courses. The researchers suggested that this may be due to a greater degree of hierarchical progression of skills in math and science, making it a better fit for instructional techniques where learners assist each other.
- Cooperative groups that were homogeneous in terms of ability outperformed heterogeneous ability groups.

This last insight is surprising given that conventional wisdom in cooperative learning is to group learners heterogeneously (Song & Grabowski, 2006; Cheng, Lam, & Chan, 2008; Hooper, Temiyakarn, & Williams, 1993; Cohen, 1994). Baer (2003) suggested that

this bias may be a result of a general belief that an important component of cooperative learning should be learning to interact with diverse groups. He pointed out that this assumption has led to a dearth of research into the differences between heterogeneous and homogeneous groups. Lou and her colleagues found that, in the comparison studies that have been done, while overall homogeneous grouping demonstrates higher effect sizes, the effects are not uniform for all students. Low ability learners appear to achieve more in heterogeneous groups, medium ability learners tend to achieve more in homogeneous groups, and the achievements of high ability learners were the same no matter which type of group was utilized. These conclusions on ability grouping were based on the findings of 12 studies.

The effects of ability grouping on learners is an area that Noreen Webb has been actively exploring for 30 years. In 1982, Webb published the results of a study where she compared the achievement of homogeneous and heterogeneous groups based on ability level in junior high mathematics classrooms over a one week unit on consumer math (Webb, 1982a). Trained observers watched the groups and noted certain behaviors: learners giving help, receiving help, asking questions, receiving no response, working alone, interacting on-task, and interacting off-task. She found that the only variable with a statistically significant relationship with achievement was asking a question and receiving no response, which was negatively correlated with achievement. This condition was more common among introverted students; they were less likely to receive answers to their questions. It was also more common in homogeneous groups. This finding suggested a superiority of heterogeneous groups, though the difference between homogeneous and heterogeneous groups was not statistically significant.

In that same year, Webb published a second study exploring the achievement of middle-ability learners in homogeneous and heterogeneous ability groups, this time utilizing a two week junior high unit on exponents and scientific notation (Webb, 1982b). She concentrated her study on three categories of student interactions: giving help, receiving help, and not receiving help. Her hypothesis was that medium-ability learners would be more likely to thrive in homogeneous groups since they would spend more time giving and receiving help than in heterogeneous groups, where higher-ability learners

would be giving most of the help and lower-ability learners receiving it. Her hypothesis was based partly on her own earlier work and partly on the work of Peterson, Janicki, and Swing (1981), who compared the performance of high-, medium-, and low-ability elementary school geometry students in two conditions: mixed-ability small group learning and large group learning. Peterson and her colleagues observed that the small group activities led to superior learning for high- and low-ability students, but that medium-ability students did slightly better in the large group condition, suggesting a curvilinear aptitude-treatment interaction for small group learning.

Webb (1982b) replicated her earlier outcome that asking a question and receiving no explanation was the single biggest factor related to post-test performance. However, in this study she also found statistically significant positive relationships between both giving and receiving explanations and achievement. One of the critical distinctions made in this study was the difference between receiving an *explanation* and receiving other types of feedback. Webb defined "receives answer without explanation" as equivalent to receiving no response at all. She did not make clear, however, what components a response needed to include in order to be considered an explanation. For example, if one student told another "the answer is 12; we do this one just like the last problem we did," is the analogous pointer explanatory enough to be considered an explanation?

In line with the elaboration perspective on cooperative learning, Webb (1982b) pointed to two mechanisms that may possibly be driving the effectiveness of elaboration as a learning tool. The first is *cognitive rehearsal*: "As a result of cognitive rehearsal, students who verbalize how to solve the problem may remember the algorithm better than students who do not" (p. 653). The second mechanism is *cognitive restructuring*. Cognitive restructuring is the process of reorganizing a mental model in order to make it easier to explain. Collectively, these mechanisms have been referred to as *cognitive tuning* (Zajonc, 1960). Given that both of these mechanisms rely on the articulation of a process, it seems likely that Webb would not have seen prompts like "do this problem just like the last problem we did" as explanations.

In terms of her hypothesis that middle-ability learners would have greater learning gains in homogeneous groups, Webb did find that "medium-ability students in uniform-

ability groups achieved more and received more explanations than medium-ability students in mixed-ability groups" (p. 652). She also found a counterintuitive outcome, that "introverted students outperformed extroverted students, but extroverted students were more likely to receive answers to questions" (p. 652). If achievement correlates with giving and receiving explanations, one would expect the opposite result. Webb, however, did not pursue introverted/extroverted studies over the long term.

In 1984, Webb published a study observing the stability of small group interaction and achievement over time. In contrast to the two studies noted above, which lasted for one and two weeks, in this study Webb tracked small group interactions in a junior high mathematics course over the course of a semester. She reaffirmed the positive relationship between giving explanations and achievement and the negative relationship between not receiving explanations and achievement. What Webb was most interested in here, though, was whether these patterns persistent longitudinally. That is, if groups stay together over a long period of time, do the early dynamics persist? Webb's data in this study indicated that group interaction effects in fact may *not* persist over time. The incidents of learners failing to receive explanations when needed dropped significantly over time, suggesting that "as students become accustomed to the small group setting, they become more aggressive in asking for help and less willing to take no response or a terminal response for an answer. Or perhaps groups become more sensitive to students' need to help over time" (p. 223).

In a summary article on small-group interaction and mathematics in 1991, Webb elaborated on the difference between explanations and other kinds of help. She defined *explanations* as

> descriptions of how to solve a problem, or part of a problem, that included some elaboration of the solution process. An example from group work on a unit on exponents and scientific notation is "Okay, look: 63,000,000 times 8,500,000. This is 63 with 6 zeroes. So, in parentheses, 63 times 10 to the sixth and then times 85 times 10 to the fifth"….Nonexplanatory help includes no elaboration of how to obtain the correct answer. Such help may consist of short answers (e.g.,

> giving the answer to a problem or exercise: "It's 10 to the -3") or procedural information such as which problems were assigned that day. (p. 368)

Still, as noted above, the boundaries of what constitutes an explanation are not well-defined, especially when moving beyond algorithmic mathematics or beyond mathematics altogether. Webb went on to suggest that there is a continuum of collaboration, with giving explanations at the high end and giving only the answer at the low end. Giving explanations correlates positively with achievement, giving only the answer correlates with negative achievement, and "giving nonelaborated information (middle of the continuum) was sometimes related to achievement" (p. 373). It is possible that a single continuum may be insufficient to describe the rich range of responses possible in social situations; Webb acknowledged that "more detailed analyses of the amount of elaboration in a responses (rather than only categorizing responses as explanations or not) are needed to clarify such a continuum and its relationship with learning" (p. 373).

In the same article, Webb (1991) commented on the benefits of receiving explanations. While explanations carry intrinsic value—people learn from verbal explanations all the time—mining this value is surprisingly difficult for learners. She argued that a number of conditions have to be in place in order for learners to gain value from explanations:

- The explanation has to address a specific need.
- The explanation has to be understandable by the learner.
- The learner actively has to apply the explanation.

Small-group learning, Webb noted, often provides an environment that supports these conditions. She also noted that an interaction does not necessarily need to be in the form of an explanation to be useful—sometimes, a simple yes or no is all that is required. Webb pointed out that most empirical results indicate that receiving explanations does not correlate with achievement, perhaps because the conditions she outlines above are difficult to put in place. While receiving explanations does not necessarily correlate with achievement, Webb stressed that "the evidence is overwhelming that receiving less

elaboration than is needed, such as asking for an explanation and being told only the correct answer, is negatively related to achievement" (p. 376).

Webb also pointed out that whether or not a help-seeker receives an explanation is not only in the hands of the help-givers. Those who ask the questions have some control. Based on the work of Wilkinson and colleagues, Webb pointed out that requests for help that are "explicit and direct are more likely to elicit explanations than requests that are vague and indirect" (p. 378).

Noted above, one condition presumed necessary by Webb for learning from explanations is that learners actively have to apply the explanation. In other words, rather than merely listening to an explanation of how to solve a problem, then moving on to a different type of problem, learners actively need to apply the explanation to a similar problem. In 1995, Webb and her colleagues studied the value of this practice by recording the behaviors of seventh-grade mathematics students working cooperatively. They concluded that learners who applied explanations did significantly better than learners who did not (Webb, Troper, & Fall, 1995).

A noteworthy feature of this 1995 study is that Webb's conception of the continuum of elaboration had evolved to include six discrete levels:

1. No response or a non-content response
2. Answer only
3. Sequence of unconnected numbers or operations
4. Numerical expression
5. Numerical rule with no verbal labels for the numbers
6. Verbally labeled explanation of how to solve part or all of the problem

In line with Webb's research, the six levels were specific to mathematics, particularly to algorithmic problems with well-structured solutions.

In 1998, after the release of the Lou metastudy, Webb and her colleagues dove deeper into the dynamics of ability and grouping by comparing the outcomes and group dynamics of junior high science students in heterogeneous and homogeneous groups (Webb, Nemer, Chizhik, & Sugrue, 1998). Their findings were that below average learners benefited significantly from the presence of high achieving learners in their

group, but that high achieving learners were better off in homogeneous groups. They found no significant differences for middle-ability students. The researchers pointed out that their findings create a "dilemma: The performance of high- and low-ability students cannot be optimized at the same time" (p. 643). They argued, though, that heterogeneous groups are a greater advantage for low ability students than they are a disadvantage for high ability students.

With this research in 1998, Webb and her colleagues developed increasingly sophisticated perspectives on several constructs. First, the researchers adapted the continuum of elaboration into a more comprehensive continuum of participation. The continuum contained three levels of participation—low, medium, and high—and included constructs absent in the continuum of elaboration, such as defending solutions and paraphrasing suggestions. Medium-level participation, for instance, was defined as including "copying someone else's suggestion, repeating it verbatim, or simply agreeing with what was said without further elaboration" (p. 619).

Another evolving construct was the researchers' clarification that the problem-solving tasks in the experiment "have well-structured solutions and answers that can be clearly explained and demonstrated" (p. 645), suggesting that the research may not generalize to more ill-structured problem solving. They noted that the more constrained the problem solving, the more likely it is that a single advanced learner will be able to single-handedly recognize and explain the solution; the less well-structured the problem solving, the more important multiple points of view become.

Finally, the researchers utilized a multidimensional view of "ability." Instead of basing ability groupings on a single metric such as a standardized test score, the researchers classified subjects based on a composite ability score of "science achievement, verbal and nonverbal reasoning ability, and vocabulary" (p. 618).

That same year, another research team studied the effects of ability grouping on high ability students, in this case elementary school student dyads (Fuchs, Fuchs, Hamlett, & Karns, 1998). Fuchs et al. found that high achieving students in homogeneous groups did much better than their heterogeneously-grouped counterparts in terms of both group interactions—the homogeneous groups were far more likely to collaborate, to react

to each other's ideas, to help each other, and so forth—and in terms of work products, though the researchers did not directly measure learning. Overall, they found a strong trend toward low achievers participating in "disappointingly menial and nonsubstantive ways" (p. 248). Low achievers in these dyads tended to act as scribes or to carry out computations as directed or to simply be "essentially nonfunctional."

Their results are inconsistent with a number of earlier studies such as Webb's work and the work of Jones and Carter (1994), who in a study similar to Fuchs et al. found that high achieving learners act as surrogate teachers, enhancing the experience for both the high achieving and low achieving learners. One possible explanation for the discrepancy is that the tasks in Fuchs and her colleagues' research were deliberately complex. In much of Webb's research, for instance, the tasks, while presumably challenging, were well-structured—one solution with a single best solution path. Fuchs and her colleagues speculated that as problem solving becomes more complex, it becomes more and more difficult for high achieving learners meaningfully to assist lower achievers, even when high achievers have training in how to give assistance, as they did in this study. This possibility may not be surprising since, as Fuchs et al. pointed out, "Recent work suggests that even trained and experienced professionals may have difficulty teaching such content to low-achieving pupils" (p. 249).

Webb and colleagues followed up by analyzing the same dataset as their 1998 study to isolate variability in performance by high ability students (Webb, Welner, & Zuniga, 2001). The researchers noted, "Whereas homogeneous groups consistently showed beneficial group functioning, only some heterogeneous groups exhibited these traits. These results show that achievement of high-ability students cannot be predicted from a simple homogeneous-heterogeneous grouping contrast and that the level of group functioning provides the key to understanding group performance" (p. 1). The researchers pointed out that their earlier analysis of the data left several puzzles, chief among them was an inability to account for different levels of achievement across group compositions. Quality of discussion did not correlate with achievement. Group composition failed to predict quantity of "high-level participation (e.g., making suggestions, paraphrasing others' suggestions)" (p. 7) for high-ability learners.

To help them better understand the efforts of high-ability learners, Webb, Welner, and Zuniga (2001) evolved their continuum for participation into multiple continua of behavior. Specifically, they looked at:

> (1) the responsiveness of the high-ability student to other students' need for help (as indicated by questions, errors, or nonverbal signals of confusion), (2) the responsiveness of the other members of the group to the high-ability student's need for help, (3) the contributions made by the high-ability student in the absence of any indication of a need for help on the part of other group members, and (4) debilitating processes occurring in the group (initiated or received by the high-ability student). (p. 12)

For categories 1 and 2, the team coded both the signals asking for help and the responses. They coded the responses "for accuracy, completeness, level of elaboration, and whether the help-giver could have provided a better or more complete response based on his or her knowledge [as demonstrated by the pre-study assessment measures]" (p. 13). They then mapped these responses to two continua of responsiveness, with the first continuum describing the content of the response, from fully-accurate and elaborated (+3) to non-responses (0) to inaccurate and elaborated (-3). The second continuum was a measure of motivation of the help-giver, from highly responsive (+3) to insulting (-3). For the other two categories the research team defined example behaviors and counted the frequency of those behaviors.

Their re-analysis of the data produced three main conclusions: high-ability learners perform well in homogeneous groups, high-ability learners sometimes perform well in heterogeneous groups, and specific types of group interactions strongly predict performance. When high-ability learners provide accurate answers to the task's questions, respond to other students who need help, and contribute answers even when no one requested them, those learners tend to do well whether the group is homogeneous or heterogeneous. The researchers characterized the presence of these conditions as "positive group functioning." When high-ability learners infrequently respond to requests for help, fail to receive help themselves when they need it, infrequently volunteer unsolicited answers, and engage in debilitating social behavior, especially trading insults,

they tend to do poorly. Webb and her colleagues suggested, “Importantly, group functioning, not group structure, predicts the outcomes for high-ability students” (p. 34). Their analysis raised the question of what causes positive group functioning.

In 2003, Webb and Mastergeorge studied the help-seeking behaviors of heterogeneous small groups of seventh-graders (2003a, 2003b). They found a pattern among subjects who did not receive high-level help during group work. These students tended to ask general questions, or to simply declare that they were confused, rather than ask specific questions. They tended to ask for help earlier, without first significantly attempting to solve the problems. In response to these vague pleas for assistance, groups tended to respond by directing those students to copy work rather than providing substantive assistance. Webb and Mastergeorge suggested this type of response is likely due either to the groups finding it too hard to provide help to someone who cannot articulate what he or she doesn’t understand, or groups perceiving these individuals as unmotivated and therefore are less likely to exert effort in their favor. The researchers suggested an alternative possibility that those individuals who exhibited poor help-seeking behaviors may be the same individuals who already have low status. The researchers discounted this possibility, though, as they analyzed academic status, gender, and ethnic background, all factors that correlate with status, and found no relationships.

Webb and Mastergeorge established in their study that specific help-seeking behaviors correlate with higher levels of helped received, which in turn correlates with higher post-test scores—the implication certainly being that students who know how to ask for help in specific ways get it, which leads to greater learning. It is theoretically possible that causality works the other way, however, though Webb and Mastergeorge do not discuss this possibility. The ability to formulate a specific question about a problem may logically demand a certain understanding of the problem—the better someone understands a problem, the more specific his or her questions can be; conversely, the less someone understands a problem to begin with, the harder it is ask specific questions. It is possible that learners who understand the problems better to begin with are likely to ask better questions or to do better on a post-test.

Worthy of note in this article is the continued evolution of Webb's tools for analyzing interactivity. Webb and Mastergeorge introduced a nine-level rubric for measuring help received. The top four levels ranked as high "because they included at least some explanation of how to obtain the numbers, the meaning of the numbers, or both" (p. 373). The rubric was specific to the particular problem focused on in the study. If a student received help that included explicit mention of particular pieces of important data in the problem and an explicit statement of the numerical rule for solving the problem, and use of particular data labels, then the researchers classified the help received as level 8, the highest level. Level 4 help meant that help-givers offered only an expression or equation without any verbal labeling. No response was level 0.

Webb and Mastergeorge also coded what subjects said or did after receiving help on a seven-level scale from "none" to explaining or reworking the problem, the highest level. Similarly, they counted how many problems subjects solved unassisted after receiving help. In fact, in this study applying help given to subsequent problems correlated more strongly with post-test performance even than advanced help-seeking behaviors.

In recent years, Webb has focused on the role of the teacher in small group learning, arguing that student help-giving behavior is an unfortunate mirror of many teachers' help-giving behavior. In a study comparing teacher help-giving behavior with peer help-giving behavior, Webb, Nemer, and Ing (2006) noted:

> Teachers tended to give unlabeled calculations, procedures, or answers instead of labeled explanations. Teachers often instructed using a recitation approach in which they assumed primary responsibility for solving the problem, having students only provide answers to discrete steps. Finally, teachers rarely encouraged students to verbalize their thinking or ask questions. Students adopting the role of help-giver showed behavior very similar to that of the teacher: doing most of the work, providing low-level help, and infrequently monitoring other students' level of understanding. The relatively passive behavior of students needing help corresponded to expectations communicated by the

> teacher about the leader as a fairly passive recipient of the teacher's transmitted knowledge. (pp. 63-64)

Whereas Webb and colleagues' rubrics for measuring group interactivity previously had been highly tailored to the specific problem to be solved, in this study they developed a five-level rubric for assessing help-giving that was more general, though still math-specific, with the highest level of help involving "at least some description of how to obtain the numbers and/or the meaning of the numbers utilized and included at least some verbal labeling of quantities" (p. 73). The researchers noted that this level of help-giving was not the gold standard for giving help in any absolute sense, only relative to the help-giving they observed, suggesting that overall the level of help provided by peers is not exemplary.

In 2008, Webb et al. studied three classrooms in which teachers received training in the elicitation of student thinking. They found that the classroom where the teacher did the most to elicit thinking, the students tended to have higher levels of discourse. However, they also observed that despite extensive training over an eight-month period, the three teachers were highly variable in their ability or willingness to utilize strategies to elicit student thinking, suggesting that refocusing teachers on behaviors that emphasize exposing and recasting cognitive processes is not trivial. In contrast, Gillies and Khan (2009) successfully utilized two days of professional development to train 28 teachers in cooperative learning questioning strategies that significantly increased the use by learners of question-asking strategies that challenged the thinking of their peers. These changes in behavior, however, did not produce significant differences on the assessment measures administered at the completion of the unit. Clearly, a great deal of work remains to disentangle the variables that impact teacher use of strategies to elicit strategic thinking and transferring this approach to learners. Webb's work here continues, as demonstrated by her recent article further establishing a correlation between teacher modeling of probing strategies and students use of them in small groups (Webb et al., 2009).

To summarize Webb's research so far, her work has advanced a number of principles related to the optimal conditions for learning in cooperative small groups:

- The act of explaining correlates with increased achievement (Webb, 1982b, 1991).
- *Receiving* explanations, on the other hand, has shown a weak and inconsistent correlation with achievement (Webb, 1982a, 1982b, 1991). Receiving accurate, understandable explanations is, naturally, more helpful than errant or confusing ones (Webb, 1991).
- Vague requests for help do not correlate with achievement, possibly because they tend to elicit low-level responses such as merely supplying the answer; more specific and persistent requests for help tend to elicit more substantial explanations driving toward conceptual understanding (Webb & Mastergeorge, 2003a, 2003b).
- Applying explanations without assistance to novel problems correlates with increased achievement (Webb, Troper, & Fall, 1995).
- Low-ability learners tend to have higher achievement when grouped heterogeneously with learners at a higher level (Webb et al., 1998). Moderate-ability learners tend to achieve more when grouped with similar-ability learners (Webb, 1982b). High-ability learners on average learn best in groups with other high ability learners, but variability is high (Webb et al., 1998; Webb, Welner, & Zuniga, 2001), suggesting it may be possible to create conditions under which heterogeneous groupings reliably lead to higher achievement for high ability learners.
- Ability grouping effects may not hold over the long term. This was a finding from early in Webb's research on this topic (Webb, 1984) but is not one that she has returned to.
- Ability grouping effects may be relative; in heterogeneous groups of low- and medium-ability learners, medium-ability learners can take on the role of a high-ability learner (Webb, 1991).
- Positive explanation-seeking behavior is often not modeled by teachers, and may even be difficult for teachers to learn (Webb, Nemer, & Ing, 2006; Webb et al., 2008, 2009).

Webb's analysis is instructive for any researcher or educator interested in effective small-group interactivity. However, her analyses hold several limitations relative to the present study. All of her work has been done on young learners, generally at the junior high level but also down into the earlier elementary grades. Her work has been done largely with group tasks centered around relatively well-defined problems. Finally, the concept of "ability grouping" is problematic in terms of transfer to "experience grouping." The sections below will analyze these limitations.

Research on cooperative learning as it relates to age. In 2000, Lou, Abrami, and Spence revisited their earlier dataset on cooperative learning to attempt to tease out, using different statistical techniques, some of the high variability they found in their 1996 meta-analysis. One new insight was that the positive effects of small group learning appear to fade with age; that is, positive effect sizes were much larger for elementary school students than for college students. This finding suggests an age effect which renders it problematic to apply research findings across grade levels. Lou, Abrami, and Spence suggested several possible reasons that would account for the negative relationship between grade level and small group effectiveness. Elementary school teachers may have more training in cooperative learning techniques than do college instructors. Teachers at higher grade levels might assume that learners at those levels need less structure than those at lower levels and therefore may not be facilitating positive interdependence or may be creating groups more prone to social loafing or other potential side-effects of combining inequality with a lack of structure. Older learners, being more used to whole class learning, may not view small group learning with positive affect. These reasons were speculative; there has been no systematic research attempting to account for the drop off in the effectiveness of small group learning for older learners. It is worthy to note here that the articles reviewed by Lou and her colleagues topped out at the college level and did not include any articles at all on small group learning as it relates to workplace training.

Of the 12 studies in the Lou et al. (1996) meta-study that compared heterogeneous and homogeneous small group learning, only one featured adult subjects (another one of the 12 studies, from 1955 and currently not practically obtainable, has an ambiguous

subject pool). The one study in this pool with adult subjects examined ability grouping among elementary education majors taking a science education class (Lawrenz & Munch, 1985). The study found an advantage for the homogeneous groups in terms of achievement but the results were not significant. The researchers did acknowledge that the groups were not structured in a way to create positive interdependence.

Further studies analyzing ability grouping in adult learners have been published since the Lou et al. meta-analysis. Watson and Marshall (1995) examined ability grouping in a college life sciences course for elementary education majors and found no significant differences between the treatment groups. The tasks carried out by the groups were loosely based on jigsaw techniques—each group member played a different role in open-ended problem-solving scenarios around ecological topics. While Watson and Marshall highlighted the contrast between their research and other ability grouping research in terms of population, observing that little ability grouping research has been done at the college level, they did not point out that their use of higher level problem-solving tasks also sets them apart from most previous ability grouping research.

Baer (2003) pointed out that generalizing cooperative learning research from young learners to adults is problematic because adults often react differently in different situations, as evidenced by prior research that group incentives appear to have a significant impact on cooperative learning in elementary schools but no impact on college-level cooperative learning. Baer compared achievement of heterogeneous and homogeneous groups in an undergraduate educational psychology course. He grouped learners at random based on quiz scores into ability groups and then had them work together on ungraded assignments such as comparing and contrasting the ideas of Piaget and Vygostky or critiquing a research design. Like Watson and Marshall, Baer's focus on relatively ill-structured problem solving creates a contrast with most existing research on ability grouping, though also like Watson and Marshall, Baer did not note the distinction. Baer kept the groups together for most of the semester and compared performance on the midterm and end of semester assessments. At the midterm (after four weeks of working together), homogeneous groups were outperforming heterogeneous groups at a borderline significant level. By the end of the course five weeks later, the homogeneous group

members were significantly outperforming their heterogeneously-grouped peers. Baer specifically compared low-, medium-, and high-achievers and found that both medium- and high-achievers achieved more in homogeneous groups and low-achievers' achievement was essentially the same across the two types of groups.

There have also been efforts to compare small learning groups of adults that have been grouped by attributes other than ability. Oetzel (2001) studied differences in grouping undergraduates in heterogeneous and homogeneous cultural groups. Schullery and Schullery (2006) grouped learners in a business communications course homogeneously and heterogeneously by personality based on the Myers-Briggs Type Indicator (MBTI) and found various effects on self-perceptions of improvement, satisfaction, and grades depending on MBTI type. Stapleton (2007) also grouped undergraduates into groups based on MBTI, this time in accounting courses, but from a different perspective. Whereas Schullery and Schullery studied how to maximize learning across members of the group, Stapleton looked instead at group performance overall—that is, he studied whether homogeneous groups outperformed heterogeneous groups on a problem-solving task and whether the groups were able to outperform the best member of the group. The ability of groups to outperform the most capable member is referred to as the *assembly effect*—Stapleton pointed out that most research indicates that groups generally cannot achieve the assembly effect. He found that heterogeneous teams outperformed homogeneous teams on a problem-solving task. However, heterogeneous groups were no more successful, statistically, at achieving the assembly effect than homogeneous groups; overall, only about a third of groups outperformed the most skilled individual. Stapleton pointed out that given the ever-growing emphasis in the business world on team building and team problem solving, "it seems imperative for secondary and postsecondary business educators to understand the factors that influence team dynamics and performance by conducting and synthesizing research to prepare students to work in teams and provide organizations with strategies to increase the capability of teams to produce quality solutions" (p. 105).

Like the present study, the Stapleton study focused on learning in the auditing/accounting field. There has been some scholarship published in recent years

related to small group learning as it specifically relates to accounting. Ravenscroft, Buckless, and Hassall published "Cooperative learning—a literature guide" (1999) in *Accounting Education*. They presented a short overview of the theoretical basis in favor of cooperative learning and an annotated bibliography of 29 books and articles about cooperative learning, of which half were specific to the accounting education field. The accounting-specific references were a mix of theoretical articles and empirical studies comparing cooperative learning techniques to non-cooperative techniques. They noted no articles comparing different cooperative grouping techniques. Lancaster and Strand (2001) likewise provided an annotated bibliography of empirical cooperative learning articles related to accounting education. They listed eight studies, generally comparing the performance of individuals working in groups to those working individually. The results of these studies were mixed, prompting Lancaster and Strand to note, "accounting education research has yet to arrive at a consensus of opinion regarding the efficacy of cooperative learning" (p. 554). They suggested that one reason this may be the case is that accounting teachers tend to use small group learning in extremely limited ways that may not demonstrate visible positive results, or may even be counterproductive. They suggested a lack of training among accounting teachers may be the culprit. In their own research connected with this review, Lancaster and Strand compared a Team-Learning Model with a traditional accounting education model but found no statistically significant differences in performance or learner attitudes.

Koppenhaver and Shrader (2003) suggested another reason why cooperative learning in accounting education has failed to achieve a critical mass of acceptance: undergraduate accounting and auditing courses tend to focus on rule-governed problems that "may lend themselves to more routine problem solving that can be performed equally well by individuals" (p. 4). Koppenhaver and Shrader's own work centered around analyzing the factors that predict smooth group functioning. Analyzing data collected from 500 undergraduates in finance courses, the researchers concluded that the biggest predictor of smooth group functioning was stability, keeping groups together over the course of the semester or longer.

The direct evidence in favor of cooperative learning for adult learners in professional-level audit or accounting training is not overwhelming. While Johnson and Johnson (2000) contended that the case in favor of cooperative learning, when it meets well-defined minimum criteria, is sound and well-documented across contexts and age groups, Lou, Abrami, and Spence's analysis (2000) raised the possibility that the effectiveness of cooperative learning fades with maturity, or at least that there are complicating factors that have not been isolated, such as differences in teacher training or different attitudes or expectations among learners. In terms of the target population of the present study, there appears to be a dearth of empirical research suggesting that cooperative learning is effective in post-college professional settings and there is conflicting evidence about whether small group learning is effective for accounting and auditing topics.

Crook (2008) suggested that very young children can work effectively in small groups but that the younger the learner, the greater the structure needed in order to overcome the natural lack of spontaneous collaboration among children—he characterized effective collaboration among young children as "strikingly rare" (p. 239). This observation suggests that maturity level matters in the small group learning equation—the open question is how those dynamics play out as learners continue to mature. Is there some minimum level of maturity beyond which it is no longer a factor in small group dynamics?

Smith (1970) demonstrated that there are inherent differences in group interactivity as a function of age. He compared the small group interactions of learners at various ages and determined that between the ages of five and twenty "positive reactions more than double in usage, questions increase six hundred fold, and negative reactions decrease to one-third of the original five year level. Attempted answers seem to increase somewhat until about age thirteen and then decrease slightly in usage" (p. 4). Smith's data indicated that growth toward adult-style interactivity in small groups is gradual, as measures like positive reactions and questions asked increase steadily as a function of age rather than plateauing. Smith's data stopped at age twenty, so it is unclear if interactivity continues to change beyond this age.

In a similar study in 1984, Dorval and Eckerman analyzed the interaction patterns in small groups of subjects at various ages, finding differences as learners aged. Dorval and Eckerman's oldest subject group consisted of college students late in their program. Like Smith, they found that the journey from child interaction to adult interaction is a gradual one: "The improvement in conversation is best described as a gradual refinement in the normative constraints on talk, rather than as an abrupt change in the prevalence of turns displaying communicative intent, as Piaget thought" (p. 66).

Smith as well as Dorval and Eckerman contended that adults have a different level of interactivity and topical coherence than do children, and that this level of interactivity develops gradually over time. However, the existing research base stops short of answering several questions germane to the generalization of the existing small group learning research base to professional training settings. Both the Smith and the Dorval and Eckerman studies focused on learners in conversational settings as opposed to learning environments. Further, it is unclear whether post-college adults are different than college age students in terms of their interaction patterns. While Dorval and Eckerman extended their research to subjects not much younger than the target population of the present study, the two populations are separated by a significant event: college graduation. Taking college courses in auditing and accounting could constitute a different experience than formal technical training in terms of immediacy and apparent applicability. In other words, studying content with an academic focus that will be evaluated by formal tests or projects for a career that is months or years away may constitute on some level a different experience than taking part in professional training designed to build skills to be applied in the immediate term and whose mastery may have a clear and present impact on performance reviews and career advancement. What is the effect of this change of context on interactivity in small group learning? Does the prospect of having to utilize the skills being taught in the class on the job immediately change the level of engagement or the types of questions asked and feedback or practice sought? Further, what is effect, no matter what the age, on small group interactivity when the skills being taught have an immediate applicability to a real world, rich context? Hooper (1992) suggested that "The effects of ability grouping may be influenced by age"

(p. 33), but restricted his analysis to young learners. Does maturity increase a high-ability learner's skill at recognizing and addressing the conceptual disconnects of lower-ability learners? Does it help lower-ability learners ask better questions and persist in finding answers? The current research base provides no insights into these questions, making generalization of the current small group research base to formal professional training problematic.

Andragogy, which attempts to tease out the differences between how adults learn and how children learn, has been called "almost certainly the best known concept in adult education" (St. Clair, 2002, p. 3). Andragogy revolves around the straightforward notion that "Teaching adults should be different if adults learn differently than children do" (Kerka, 2002, p. 3). Andragogy is built around six key assumptions about how adults learn differently than children (St. Clair, 2002). Adults:

- Are self-directed.
- Have a much greater pool of experience to draw on.
- Are oriented toward learning to solve real world problems.
- See education as a path to competence.
- Need to know the reason for learning something.
- Are largely motivated by internal forces such as self-esteem.

Ideally for the purposes of the present study, andragogy would elucidate principles that would help predict how the body of cooperative learning literature would translate to adult, professional learners in technical training situations. For instance, andragogy suggests that adults, having identified their needs and interests, are self-directed (Robles, 1998) and therefore should need less structure in small group learning.

Andragogy, however, has come under considerable attack from various camps. St. Clair (2002), in musing on the question "Is andragogy a theory for the 21st century?" offered:

> Addressing this question involves considering three uses of andragogy—as a learning theory, as a guide to teaching, and as a foundational theory of adult education. In the first case, there appears to be general consensus that andragogy does not perform any of the functions of an adult learning theory in terms of

> explaining how and why people learn. It is almost certain that andragogy has been both undermined and superceded as a learning theory. (p. 4)

Critics contend that it is a vast oversimplification to suggest that human suddenly become self-directed when they reach adulthood, or that adults at any age are self-directed in all their learning-related activities—such a view of adulthood ignores relationships of power and who gets to decide who needs to learn what (Kerka, 2002). Further, andragogy has been criticized for suggesting by inference that children are not self-directed and are only motivated by external forces. Researchers such as Kerka see the principles of andragogy as suggesting continua on which adults tend to fall more toward one side than children do, acknowledging that the diversity of adults renders those continua powerless to predict the needs or behavior of any individual adult or small group of adults.

Taylor, Marienau, and Fiddler (2000) focused on experience as the key differentiator between children and adults: "The major difference between adults and younger learners is the wealth of their experience. They have seen, been, and done….These experiences are valuable—we would claim *essential*—contributions to the learning process" (p. 7). Even in a small group of learners, there may be a significant range of real world experience with the subject matter—the effective sharing of which may be critical to the success of the small group—yet heterogeneity of *experience* has not been among the foci of research into cooperative learning.

Research on cooperative learning as it relates to experience. Hooper, Temiyakarn, & Williams (1993) asserted that researchers need to expand the range of factors they explore when considering cooperative learning, especially as they relate to group composition. They acknowledged efforts to research factors such as gender, ability, and personality characteristics, but suggested researchers should branch out to explore additional grouping criteria. Lou et al. (1996) labeled it "unfortunate" that "a scant number of studies exist to integrate findings on group composition criteria other than relative ability or prior achievement" (p. 427). The emphasis on ability grouping perhaps is not surprising given McAuliffe and Dembo's finding (1994) that achievement is the single biggest status marker in cooperative learning groups; the individual that a small learning group perceives as having the highest ability is the one with the most influence.

A learner's incoming *ability* is typically defined in ability-grouping studies as achievement on a pre-test or previously-achieved grades, though it is sometimes augmented by measures like scores on tests of reasoning ability (e.g., Webb, Welner, & Zuniga, 2001). These scores are then used to group learners by ability. One situation that might warrant a different pre-assessment of ability is when formal tests of ability are not practical. While testing is just a part of kindergarten through university life, it may not always be desirable or feasible at the professional training level. Testing takes time, which in a business setting can bring significant scrutiny from a cost-benefit perspective. In an audit training setting, for example, time spent in the classroom is time spent not charging to projects which in turn are charged to customers. In the professional experience of this researcher, time spent in training is often a practical compromise, not necessarily determined by the length of time the course developers feel would be optimal for teaching the content; course length and scope are sometimes determined before the course is designed or objectives clarified. Compromise is often the order of the day. This does not mean that a pre-test or other formal determination of ability is impossible; it does, however, raise the practical desirability of a quicker assessment. Domain-specific experience may be a reasonable analog. A short round of introductions is not uncommon in formal training events. By asking appropriate questions during these introductions, instructors may be able to gauge quickly relevant experience to ensure either that all the most experienced learners are together or that they are distributed throughout the small groups.

Professional training may at times also present situations where traditional assessments of ability may not be the optimal choice for grouping even if some time was set aside. Written tests, by their nature, tend to reflect well-structured problem solving. As Jonassen, Strobel, and Lee (2006) put it, "Word problems possess knowable, correct solutions that are achieved by applying preferred solution methods; and they apply a limited number of regular rules and principles that are organized in a predictive and prescriptive arrangement" (p. 139). Workplace problems, however, are often ill-structured—complex problems that possess multiple goals and potential solution paths, non-technical success standards, multiple constraints and unanticipated problems, and

essential knowledge distributed among multiple team members. There is evidence that the ability to solve well-structured problems does not readily transfer to solving ill-structured problems (Shin, Jonassen, & McGee, 2003), suggesting that test performance may not be the most appropriate criteria for grouping learners in all professional training situations. Domain-specific professional experience—placing learners into *experience* groups rather than *ability* groups—may at times be a more appropriate choice.

Problem-solving situations do not necessarily fall neatly into either the well-structured or ill-structured categories. Jonassen (1997) described the split more as a continuum. Ill-structured problems tend to feature characteristics such as:

- One or more elements of the problem are unknown or at least uncertain.
- The goals are vaguely defined. There are multiple possible solutions and significant room for disagreement about the criteria for a successful solution.
- Multiple constraints are in play, some unstated.
- Cases cannot be solved by the application of a small number of rules or principles.
- Learners have to utilize significant judgment and defend their solutions.

At the far end of the continuum, extremely ill-structured problems have been characterized as *wicked problems*—those intractable social policy problems of immense scope that lack agreement around the nature of the problems (Rittel & Webber, 1973).

Ill-structured problem solving is a core component of many professions. Jonassen, Strobel, & Lee (2006) highlighted the prevalence of ill-structured problems in engineering and contrasted those problems with the well-structured problems typically featured in engineering programs at universities. Both Jonassen (1997) and Ertmer et al. (2008) upheld instructional design as a profession replete with ill-structured problem solving. Ertmer et al. used the example of converting a workshop from a classroom to an online format, a common instructional design task that offers many potential solution paths but no single correct answer. The success of proposed solutions are based not only on upholding instructional design principles, but also on successfully balancing real world constraints.

The ability to solve typical domain problems consistently is the mark of an expert in a field. Expertise consists of two major components, "abstract knowledge and practical experiences" (Ertmer et al., 2008, p. 72; see also Levitt & Dubner, 2009), both of which are necessary in order for experts to solve a wide variety of problems in their domain. There is evidence that experts are able to access and apply the appropriate facts and experiences to ill-structured problems often through the recall of practical experiences rather than by directly recalling abstract knowledge (Kolodner, 1997). This position forms the basis of the field of case-based reasoning (CBR).

Ill-structured problem solving and case-based reasoning. As Kolodner (1997) defined it:

> CBR focuses on reasoning that is based on previous experience. A previous experience might suggest a solution to a new problem or a way of interpreting a situation, may warn of a problem that will arise, or may allow the potential effects of a proposed solution to be predicted—all types of inferences necessary for addressing the kinds of ill-defined or complex problems that come our way in the workplace, in school, and at home. (p. 58)

From a CBR point of view, novices gain expertise by encoding, accumulating, and retrieving cases into a functioning case library in long term memory. This process depends on acquiring a bank of cases encoded in such a way that they can be usefully retrieved at the time of need; generally, this means encoding cases based on the deep structure, the underlying principles and concepts, rather than on surface features (Jonassen & Hernandez-Serrano, 2002). Proficiency in problem solving in a domain involves invoking a blend of practical experience with abstract knowledge (Ertmer et al., 2008; Genberg, 1992; Levitt & Dubner, 2009)

Cases can be accumulated by learners for use in problem solving in at least three ways (Kolodner, 1997). One is direct experience solving real world problems. A second way is through simulation. A third way is to make an external case library available during problem solving, one made up of cases told by individuals with problem-solving experience in the domain. Kolodner noted, "Such cases, even though they have been experienced by others, can suggest issues to focus on and solutions to problems, warn of

potential pitfalls, support projection of the effects of a chosen solution, and so on, facilitating solution of more complex problems than students could solve by themselves" (p. 62). In effect, learners can utilize the case library to have vicarious problem-solving experiences that they can leverage similarly to actual experiences—as Jonassen and Hernandez-Serrano (2002) put it, "The memory structures employed while understanding and applying these stories are similar to those employed if the learner had originated the solution…in other words, understanding is functionally equivalent to experiencing" (p. 74; see also Ferguson, Bareiss, Birnbaum, & Osgood, 1992).

One advantage of cases as a teaching tool is their narrative structure—each one tells the story of someone with a goal and how they met that goal, or failed to meet it. Stories, noted Jonassen and Hernandez-Serrano (2002), "require less cognitive effort than exposition because of the narrative form of the framing experience" (p. 66). They contended that swapping stories is how people most naturally convey problem-solving information. Referencing the ethnographic work of Julian Orr in studying the discourse of technicians, Jonassen and Hernandez-Serrano continued:

> Orr found that these practitioners employed storytelling for framing and dealing with problems. Narrative was used for explaining catastrophes; for understanding, explaining, and arriving at diagnoses; for teaching and learning new methods; for dealing with uncertainty; for changing perspectives on problems; for warning about failures; for providing solutions; for expanding the problem space; for finding causes to problems; for illustrating a point; for challenging a fellow technician; for building confidence as problem solvers; and for anticipating future problems. (p. 68)

Jonassen and Hernandez-Serrano noted that another important function of narratives is that they help initiate new members into an organization. Stories are used by mentors to transfer memories and reflections.

CBR is focused on building and deploying case libraries for use in instruction. While libraries can be built by interviewing experienced problem solvers, too much experience may actually work against the creation of effective cases as high level experts may have difficulty articulating their underlying domain knowledge (Jonassen &

Hernandez-Serrano, 2002). There may well be an parabolic relationship between level of expertise and effectiveness of cases—the case generators have to have enough experience to convey cases that are both accurate and insightful but not so much experience that they can no longer put themselves in the position of novices to understand their knowledge gaps and misconceptions.

CBR is centered around case libraries in part to take the burden off the instructor to relate a wide variety of narratives at the right level and because learners are unlikely to have sufficient experience to be able to help each other (Jonassen & Hernandez-Serrano, 2002). In many traditional educational settings, it may be reasonable to assume that learners will not have the practical real world expertise to be able to convey effective cases to each other. However, in courses that include learners with a wide range of experience, more experienced learners can theoretically serve as adjunct case libraries for inexperienced peers—Ge and Land (2004) characterized premade supports as *hard scaffolds* and human supports as *soft scaffolds*. This line of reasoning suggests that heterogeneous experience groups may be superior to homogeneous experience groups in terms of assisting novice learners to achieve greater learning gains. Applying the principles of CBR, this positive effect is contingent on several factors. First, experienced learners have to have the right kinds of experiences to convey. They themselves have to have had sufficient meaningful problem-solving experiences for them to ensure understanding of the domain knowledge—describing an incorrect or woefully inefficient way to solve a problem is presumably not helpful to novices. Second, the experienced learners have to actually convey narratives of those experiences, either spontaneously or when prompted. Third, they have to convey them in a way that helps novices encode the cases in a way that makes them retrievable at the right time. This involves conveying the deep structure—giving the *whys* in addition to the *hows*—so learners can generalize the vicarious experience to new cases. Fourth, they have to convey the case at the right time, at times when novices can apply the cases to problem-solving tasks.

In short, in order for experienced learners to serve as case libraries, a number of factors have to fall into place. Hard scaffolds eliminate the risk of variability. However, experienced learners conveying cases to each other and to novices holds one significant

advantage over premade libraries: their narratives are dynamic instead of static—they can respond to questions, clarify points, and elaborate spontaneously. In studies of how experienced banking consultants conveyed information to novice consultants, Ferguson, Bareiss, Birnbaum, and Osgood (1992) noted the importance of the learner being in control, asking for information and clarifications. The researchers referred to this phenomenon as *conversational coherence*.

Another logical real world advantage of experienced learners as case libraries is reduced development costs; collecting cases takes development effort, often a scarce commodity. Experienced learners serving as soft scaffolds may thus reduce instructional development efforts. These savings would negatively impact quality, however, if experienced learners were unable to serve effectively in this capacity. Generally, even the most experienced learners in a classroom will still be relatively inexperienced; if they convey information via cases that is wrong or if they convey only the surface features, the effect may be counterproductive to learning.

Ill-structured problem solving and the auditing field. The more ill-structured the problem solving, the more important the role of experience and case-based reasoning (Shin, Jonassen, & McGee, 2003). Therefore, an important question to consider for the present research is: How ill-structured are the problems that new auditors are typically asked to solve? The more ill-structured the problem solving for new auditors, the more critical the role of prior experience. Messier, Glover, and Prawitt (2006) define auditing this way:

> Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users. (p. 12)

Starting Your Career teaches that auditing is the act of opining on the financial statements of an entity in regard to the question of whether they are free from material misstatements. Auditors provide reasonable assurance to readers of financial statements that they are substantially correct. Material misstatements can arise from two sources: error and fraud. In order to establish that the financial statements are free from material

errors and fraud, auditors have to find and critically examine evidence. Modern corporations are extremely complex entities; while there are myriad laws and regulations in place to define what financial practices are acceptable and which are not, they cannot begin to guide unambiguously every financial business decision. Therefore, decision makers in corporations frequently have to interpret the rules and regulations. Auditors need to evaluate whether these interpretations are reasonable, which requires judgment.

Further, *Starting Your Career* teaches that in order to exercise judgment successfully, auditors need to understand an entity's financial processes, evaluate which processes are most prone to material errors, and collect evidence to verify that no material errors are present in the period covered by the financial statements. When it comes to fraud, auditors have to evaluate who has access to, and motivation and rationalization for stealing, significant funds. It should also be noted that auditors do not provide *absolute* assurance, but rather *reasonable* assurance—establishing what is reasonable takes judgment. In fact, the professional auditing standards define *reasonable assurance* in terms of judgment: "While exercising due professional case, the auditor must plan and perform the audit to obtain sufficient appropriate audit evidence so that audit risk will be limited to a low level that is, in his or her professional judgment, appropriate for expressing an opinion on the financial statements" (AICPA Statement on Auditing Standards No. 104.10, 2006).

Further, the success of an audit is not only measured by how successfully the engagement team verifies that the financial statements are free from material misstatements. Its success, from the perspective of the audit firm, is also measured by whether the audit engagement was completed profitably.

Auditing firms create extensive processes, procedures, and forms to guide audit engagement teams, but these vary from firm to firm and in any case are only guidelines. Audit teams have to perform whatever measures are necessary to provide reasonable assurance, and that can vary significantly from engagement to engagement. Auditing itself, then, can reasonably be considered to engage ill-structured problem-solving. The ill-structured nature of the whole task, however, does not automatically mean that novice auditors are tackling ill-structured sub-tasks.

A difficulty in determining whether novice auditors perform ill-structured problem solving is that the continuum of well- to ill-structured problem solving is itself ill-defined. While David Jonassen has laid out characteristics of well- and ill-structured problem solving, no validated typography exists to help instructional designers or researchers to categorize problems on the continuum. This lack of specificity makes it difficult to compare research across contexts. If problems typically faced by beginning auditors strongly meet some of the criteria for ill-structured problems but only weakly meet others, does that mean that the problems should not be considered ill-structured? Absent a formal, rigorous typology, problem-solving activities engaged in by new auditors will be described below in terms of Jonassen's criteria (1997) distinguishing well- and ill-structured problem solving.

Given problem elements. In well-structured problem solving, all elements of the problem are known and are given as part of the problem statement. In ill-structured problem solving, one or more problem elements are either not known or are not known with a specific degree of confidence. In terms of auditing, even new auditors have to look critically at the evidence they have and decide if it is enough or if they need to acquire more. In the course featured in the present study, a number of the simulations include elements where the starting set of information proves to be inadequate to build sufficient audit evidence and learners have to figure out what they need and ask for additional information.

Goal definition. In well-structured problem solving, the goals are given and well-defined. In ill-structured problem solving, the goals are unclear and open to definition. A novice auditor's tasks vary on this continuum. On one end of the spectrum, even new auditors are expected to come to the first day of an audit engagement having done some preliminary research on the client and the client's industry. However, this research is open-ended. There is not any particular knowledge they are supposed to have, just enough information so they can ask intelligent questions during the audit. Likewise, before auditing the cash accounts on the financial statements, auditors are expected to take time to prepare utilizing last year's statements. Again, the preparation is not well-defined. Auditors look for trends, unusual events, insights from previous years, and so

forth that can help them find elements in the cash section worthy of greater scrutiny. A typical task in the course scrutinized in the present study would be to give learners a set of workpapers from last year's audit and ask learners to create a list of conclusions reached that might be relevant to this year's workpapers. On the other end of the spectrum are common, relatively well-defined tasks, such as "footing" a client's statements by validating their spreadsheet formulas.

Constraints. In well-structured problem solving, all constraints are given as part of the problem. In ill-structured problem solving, the constraints are initially unstated and are only identified or discovered during the act of problem solving. In terms of auditing, inexperienced auditors are often unaware of constraints on audit engagements, including hours budgeted for specific tasks and constraints about how they can interact with the client. One of the goals of *Starting Your Career* is to make learners aware of the various kinds of constraints that can affect them depending on the audit engagement and to ensure they seek out that information in a timely fashion.

Solution paths. Solving well-structured problems involves the application of a limited number of rules and principles to follow a single solution path toward the single correct answer. Ill-structured problems can be solved via multiple solution paths—or, there may be no adequate solution paths at all. In terms of auditing, modern audits even of small companies are highly complex with lots of moving parts and rules and principles in play, not all of them known or clear. While for many audit tasks, especially for novice auditors, there are a limited number of paths to achieving the goal, for others there really are a multiplicity of paths. For instance, interviewing clients for information can be done multiple ways—and indeed sometimes there effectively is no answer as the auditor simply will not be able to get the client to share the needed information. In these cases, the only recourse is to escalate the problem to the individual "in-charging" the engagement.

Definition of success. Well-structured problems have solutions that are clearly defined, making it possible to distinguish correct from incorrect answers reliably. Solutions to ill-structured problems are evaluated by multiple, possibly conflicting, criteria, such as balancing quality and efficiency. The work of novice auditors is

continually monitored by more experienced auditors who are looking for work that is complete and compelling but also done efficiently. Likewise, auditors are evaluated against not only how they audit the client but also how they interact with clients—and with the audit engagement team.

Prototypical cases. With well-structured problems, it is possible to create a single prototypical case that can serve as a recipe for solving all cases in that class of problems. No prototypical cases are possible for ill-structured problems because the cases vary too significantly or the variables interact in highly complex ways. One of the core messages of the present course is that auditors should not blindly apply prototypical models to their audit work. Generally, these prototypical models take the form of last year's audit notes and workpapers. Novice auditors are often tempted to repeat the procedures that the audit team did last year without critically considering whether anything has changed that would make different procedures more appropriate, or even whether the procedures performed last year were the most appropriate in the first place.

Solution defense. Solutions to well-structured problems can stand on their own as either right or wrong. Solutions to ill-structured problems must be defended. As above, the only output of an audit is an opinion and a set of workpapers justifying that opinion. Every conclusion an auditor reaches needs to be defended in the workpaper documentation. Creating persuasive documentation is one focus of the course.

Ge and Land (2004) call the final criteria—that solutions must be defended—the "characterizing" process in ill-structured problem solving as the problem solver "must be able to make claims about the probably effects of events, objects, or phenomena on others, warrant those claims, and back them up with supportive statements, facts, or conjectures" (p. 12). Ge and Land noted that this area of problem solving is especially troublesome for learners—they tend to confound arguments with opinions, forgo evidence, and leave arguments underdeveloped. Likewise, Shin, Jonassen, and McGee (2003) found, in an astronomy simulation, that justification skills, along with domain knowledge and metacognitive skills, to be the best predictors of ill-structured problem performance.

In terms of auditing, justification of conclusions reached is critically important for all auditors, including new auditors. The AICPA Audit Standards Board Statement on Auditing Standards (SAS) No. 103 (2005) lays out the professional standards for audit documentation:

> The auditor should prepare audit documentation that enables an experienced auditor, having no previous connection to the audit, to understand:
>
> a. The nature, timing, and extent of auditing procedures performed to comply with SASs and applicable legal and regulatory requirements;
> b. The results of the audit procedures performed and the audit evidence obtained;
> c. The conclusions reached on significant matters; and
> d. That the accounting records agree or reconcile with the audited financial statements or other audited information. (No. 103.10)

SAS No. 103 goes on to note that it is "neither necessary nor practicable to document every matter the auditor considers during the audit" (SAS No. 103.11) and then outlines the factors that auditors should consider when deciding what to document, including the nature of the auditing procedures and the extent of judgment involved. SAS No. 103 is 35 sections long, not including the extensive appendix, and provides significant scaffolding to auditors in justifying their positions, in part meeting the calls from Ge and Land and others calling for scaffolding in helping learners construct arguments. Beyond that, however, the emphasis in the professional auditing standards on documentation lends weight to the argument that ill-structured problem solving is an integral part of auditing at all levels.

That said, auditing clearly does not approach the wicked problem end of the problem-solving scale. There is not really a blank canvas that invites wildly divergent thinking. Two audit engagements of similar companies in similar industries are going to have a great deal of overlap in the procedures performed. While professional standards bodies go out of their way to emphasize the importance of auditor judgment, audit firms still create detailed methodologies in order to create consistency across engagements. Still, even small differences in an audit can make the difference between uncovering

material misstatements and missing them, and auditors have to be thinking critically in order to find the places where a small deviation in audit procedures can make a big difference in the outcome of the audit. Since so much of auditing is about defending professional judgment, the field cannot be characterized as well-structured. However, there is enough consistency across audits that it also cannot be said to be at the far end of ill-structured. Absent a more rigorous typology, it may not be possible to categorize problem solving in a course more precisely.

Research into cooperative learning and satisfaction. Oetzel (2001) characterized satisfaction as an important consideration for any small group situation because it can affect the cohesiveness of the group, which can affect not only present performance but also future participation. However, of the six studies discussed earlier that studied effectiveness as a function of group composition for adults, four did not track satisfaction as a variable of interest (Lawrenz & Munch, 1985; Watson & Marshall, 1995; Baer, 2003; Stapleton, 2007). Of the remaining two studies, Oetzel (2001) concluded that satisfaction is positively correlated with equivalent levels of participant and respect, but that equality of participation and respect do not correlate with performance. He did not link satisfaction to particular types of group composition. Schullery and Schullery (2006) also found a complex relationship between satisfaction and performance, suggesting that grouping learners for diversity in terms of "argumentativeness" improves satisfaction but decreases performance.

A lack of positive correlation between satisfaction and performance does not necessarily decrease the importance of satisfaction as a variable of interest when studying the effectiveness of a particular group type, particularly with adult learners. Measuring satisfaction, synonymous with level one of Kirkpatrick's four-level evaluation framework (Kirkpatrick & Kirkpatrick, 2006), is extremely common among businesses that conduct internal training—far more common than higher levels of evaluation (Bassi, Benson, & Cheney, 1996). In these businesses, *satisfaction* by default becomes synonymous with *effective*. If one type of cooperative learning grouping raises satisfaction without harming other measures of effectiveness, that would be a positive finding for many real world situations.

Further, there is evidence that job satisfaction correlates with job performance. A meta-study conducted by Judge, Thoresen, Bono, and Patton (2001) found a correlation of .30—more satisfied employees tend to perform at a higher level, though these results are disputed (Bowling, 2007). Satisfaction with training in turn has been correlated with overall job satisfaction (Chiang, Back, & Canter, 2005). Thus, it is possible that increased satisfaction with training can impact overall job satisfaction, which in turn can increase job performance in ways that go beyond the skills taught in training. From an organizational perspective, this makes satisfaction a compelling element of an effective course.

Summary and Framework for the Present Study

The present study views cooperative learning from a cognitive perspective, particularly an elaboration perspective (O'Donnell & O'Kelly, 1994; Slavin, 1996). This perspective assumes that the act of articulating one's cognitive model helps a learner to refine, reinforce, and enhance the model, leading to more robust learning. It further assumes that received explanations are valuable to the recipient in terms of increasing understanding, but only if the explanations meet certain criteria, such as being understandable to the help-seeker and, where appropriate, focusing on the underlying problem structure. Therefore, cooperative learning opportunities that generate a significant number of high quality exchanges of explanations should lead to greater learning than those that generate fewer exchanges or more surface level dialog.

Further, the present study embraces the assertion of Merrill and Gilbert (2008) that the best starting point for generating high quality peer dialog is by presenting learning groups with a well-constructed sequential iteration of meaningful problems. The implication from Merrill and Gilbert's work is that the best path to cooperative learning is not the grafting of a cooperative learning structure onto existing traditional instruction, but rather the best path is to start with challenging small groups to solve meaningful problems, adding scaffolding as necessary. Enough scaffolding should be applied to maintain an appropriate level of challenge and dialog but not so much to prevent learners from having to think in sufficient depth to organize the materials collaboratively (Cohen, 1994).

The present study will not challenge the conventional wisdom that cooperative learning can enhance learning over individualized or competitive learning regardless of academic discipline and maturity of learners (Johnson & Johnson, 2000), even though the evidence in favor of cooperative learning seems to decrease as a function of age (Lou, Abrami, & Spence, 2000) and empirical studies of cooperative learning in post-college formal learning settings appear to be non-existent. However, the present study does challenge the default notion present in many cooperative learning texts that learners should be grouped heterogeneously by ability in cooperative learning groups.

The most comprehensive meta-analysis available regarding cooperative learning (Lou et al., 1996) provides evidence that low ability learners achieve more in heterogeneous groups, medium ability learners achieve more in homogeneous groups, and high ability learners exhibit no significant difference, with the balance across all learners tipped toward homogeneous grouping. From the perspective of the present study, however, the meta-analysis findings, based on 12 comparison studies, are problematic in a few ways. First, none of the studies compare ability groupings beyond the college level, and only one of the 12 was even at the college level. Second, the comparison studies are not broken out by level of problem solving. There is evidence that adults inherently interact differently in groups than adolescents (Smith, 1970; Dorval and Eckerman, 1984). There is also evidence that groups assembled to solve well-structured problems may act differently than groups assembled to solve ill-structured problems (Fuchs, Fuchs, Hamlett, & Karns, 1998).

A third limitation of the existing body of ability grouping research is that *ability* as it is typically defined in research studies—generally as higher scores on pre-tests or similar academic metrics—may not be the optimal measure of ability for adult professionals. Especially in a profession such as auditing, where by their nature many of the problems are ill-structured, experience may be a superior measure of ability than knowledge or performance solving well-structured problems (Jonassen, Strobel, & Lee, 2006). Experience may also be easier to gauge quickly and without the use of formal pre-testing, a big advantage in technical training courses where time is short and precious.

This study will compare, in a new hire auditing course in a medium-large financial services firm, the learning gains of participants grouped homogeneously by experience level with those grouped heterogeneously. The course was recently redesigned by an internal team, of which the present author was a member, to feature a logical progression of simulated audit problems built within a basic architecture of modeling, either by the teacher or via prerecorded media, followed by scaffolded practice and small group simulations. The problem solving taking place in the class reflects the typical range of ill-structured and well-structured problems that are the core of the job for first year auditors.

Based on the extensive work of Noreen Webb, achievement should correlate with:

- Giving detailed conceptual explanations.
- Specific and persistent requests for help.
- Applying explanations to novel problems unassisted.
- A lack of debilitating, distracting behaviors.

The present study will adopt Webb's qualitative techniques for measuring these behaviors during problem-solving dialog. A pre-test/post-test methodology will be utilized to measure quantitative learning gains. A standard survey already utilized by the host firm will measure satisfaction.

Based on the Lou et al. (1996) meta-analysis findings, the researcher predicts that homogeneous experience groups will feature more detailed conceptual explanations and more specific and persistent requests for help and thus greater learning gains. There is currently no basis in the literature to predict which type of group will have greater satisfaction.

Chapter 3: Method

This chapter describes the research method for the present study. Included are descriptions of the participants, materials, experimental design and treatments, data analysis, and procedures. The methods described were designed to answer this research question:

When professional, adult learners in problem-centered technical training courses are grouped heterogeneously and homogeneously by content experience, what is the effect on course effectiveness as measured by:

1. Learning gains
2. Problem-solving discourse
3. Learner satisfaction

Participants

This research took place at a medium-large financial services firm headquartered in the American Midwest. A total of 46 new auditors were hired by the firm in late 2008; all took part in a one-week course entitled *Starting Your Career* and all voluntarily participated in this study. Nineteen of the 46 learners were female (41%). The course is required for all new audit hires in the firm, so participants in this course came from several points across the country.

The participants varied in previous audit experience. Twenty-seven participants had never participated in an audit engagement, though about half of those participants had participated in one or more related financial services engagements such as tax. The remaining 19 learners had participated in at least one audit engagement, either as an intern or in between the time they were hired and the beginning of *Starting Your Career* in early January, 2009. Two learners had been a part of more than ten engagements, with one having been on 11 engagements and the other on 15. The distribution of audit experience among learners with experience is shown in the graph below.

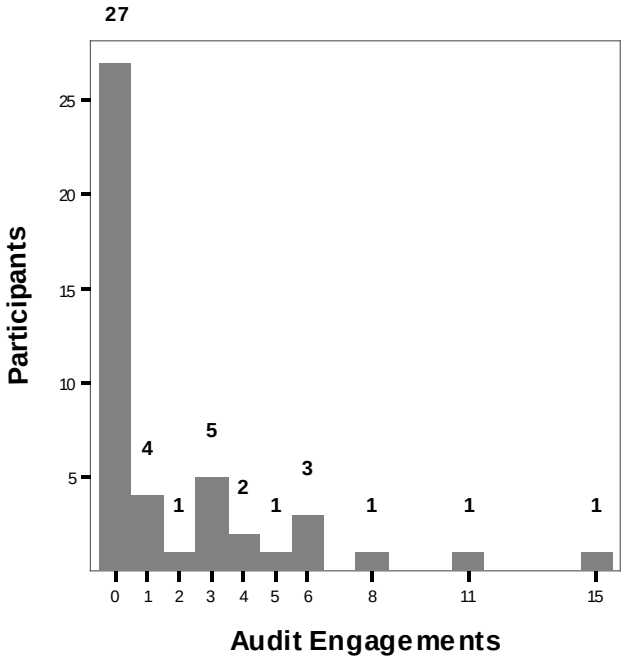

Figure 1. Number of Participants with Different Levels of Audit Experience

The vast majority of learners were taking *Starting Your Career* for the first time; however, because three of the learners had previously completed internships with the firm, they had previously taken the old version of the course.

A typical profile of a new auditor is that he or she has recently completed college with a degree in accounting. The vast majority of participants in this study fit this profile; however, a small number of the participants came from non-traditional backgrounds and were embarking on auditing as a second career. Many new auditors serve in an internship with an audit firm over the summer after they graduate and then begin working full time in the field either in the fall or the winter afterwards. Auditors generally work in the field for a period of years before passing the exams to become a Certified Public Accountant (CPA). None of the participants in this study had passed the CPA exam at the time of the study.

Materials

The materials used in this study consisted of a week long course guided by a detailed leader guide and participant guide, a paper-based pre-test and post-test, a

cooperative discourse instrument, semi-structured participant interviews, and a formal course evaluation.

The course. *Starting Your Career* is a week long course co-taught by two instructors. The course was designed to meet the objectives described in the following table.

Table 2. Course Objectives

Upon completing *Starting Your Career*, participants will be able to:

• Describe the basic purposes of an audit engagement.
• Describe one or more ways that real world users of financial statements utilize the statements and ways that an audit can impact those uses.
• Describe the purpose and output of each basic stage of an audit: planning, fieldwork, and reporting/wrap-up.
• Identify, given a description of a hypothetical company, realistic potential errors and fraud that could materially impact the company.
• Describe the basic business process flow: transaction, journal entry, sub-ledger, general ledger, trial balance, financial statement.
• Discriminate processes from controls.
• Describe relevant audit program steps and how they relate to reducing the risk of material misstatement.
• Describe the role of an associate in an audit.
• Recognize when to research an issue and identify appropriate Firm resources to research it.
• Navigate CaseWare, including:
  o Import and map trial balances.
  o Embed and link files.
• Write a memo that describes procedures performed in sufficient

Upon completing *Starting Your Career*, participants will be able to:
detail that another auditor could replicate those procedures and document work performed, including appropriate use of tickmarks and references. • Interview clients to secure information in a professional manner. • Recognize when to take notes. • Behave in line with the Firm culture of professionalism and etiquette. • Complete the basic audit program steps. • Evaluate audit evidence demonstrating appropriate professional skepticism. • Navigate Lotus Notes to find Firm's tools and manuals.

The instructional architecture of the course was inspired initially by the cognitive apprenticeship model (cf. Collins, Brown, and Holum, 1991; Dennen & Burner, 2008). Cognitive apprenticeship provides a blueprint for building cognitive strategies that integrate skills and knowledge in order to accomplish meaningful tasks. Similar to an apprenticeship in the traditional sense, cognitive apprenticeship relies on modeling, scaffolding, and fading to build strategy use. Unlike the physical skills generally taught with traditional apprenticeships, however, cognitive skills are not plainly visible, so the cognitive apprenticeship model focuses on exposing cognitive strategies.

The cognitive apprenticeship model revolves around five key instructional steps: modeling, coaching, articulation, reflection, and exploration (Collins, Brown, and Holum, 1991). The design team addressed these steps in various ways, as described below.

Modeling. In order for learners to understand what cognitive strategies they are expected to learn and utilize, the strategies have to be modeled out loud. Major strategies taught in *Starting Your Career* were always modeled for learners before asking them to practice. This modeling took two forms. For some sections, the modeling was captured as pre-recorded video. For instance, in a module about interviewing clients to gain

information, the design team scripted and videotaped simulated client interviews that contrasted positive and negative examples. In other sections, the instructors were given guidance in how to model the skill first while thinking aloud to expose the critical thought processes. During modeling, participants were directed to their participant guides for scaffolded note-taking; for example, participants might be asked to write down the major steps the instructor took along with a note about how that step fits with the larger goals.

Coaching and articulation. After being exposed to appropriate models, learners should then be supported during practice activities with appropriate scaffolding. During these activities, learners should be asked to articulate the strategies they use. Modeling in *Starting Your Career* was generally followed up with a scaffolded practice phase. This phase often took the form of analysis of written case studies. For example, participants might be asked to read the case study and then in their small groups determine why the auditor took certain steps, or they might be asked to flag steps that were not efficient nor useful.

Reflection. Learners should be asked to perform self-analysis and assessment. To help learners assess their progress in *Starting Your Career*, many modules were punctuated with "check your understanding" questions. Less formally, groups were directed and reminded to check in with each other to ensure no one was being left behind.

Exploration. Learners should eventually enter a mode of problem solving on their own. In *Starting Your Career,* the third major phase after modeling and scaffolded practice would be for participants to practice the skill directly without formal scaffolding. In their small groups, learners would practice skills like interviewing simulated clients or researching issues using the firm's research tools. In addition, whereas most of the week-long course was devoted to auditing cash accounts, a common task for first year auditors, the final day was devoted to a less scaffolded simulation around auditing accounts payable, a related type of account. This gave learners the opportunity to explore how their strategies generalize to different accounts.

Ultimately, the design team ended up focusing more on modeling, coaching, articulation, and exploration than on reflection. From a cognitive apprenticeship

perspective, this was a weakness in the design. Further, while learners had significant opportunities for unscaffolded practice, opportunities for learners to choose areas of interest to explore, which Collins, Brown, and Holum (1991) suggested is the natural next step of problem solving, were limited.

Pre- and post-test. Formal testing of learners is rare in this organization and there are no industry standard tests available to gauge a new auditor's readiness to enter the field. Therefore, pre- and post-tests were created by the design team by pairing an instructional designer (the author of this study) with a subject matter expert.

A pre-test of 14 items, a combination of multiple choice, fill in, and short answer was constructed to reflect the key course objectives. The test was reviewed by a second subject matter expert and then pilot tested with two first-year auditors who had already completed *Starting Your Career*. The team revised the test based on feedback from the pilot testers and subject matter expert, and then developed a post-test that closely paralleled the pre-test but that varied critical attributes and variables where feasible.

Cooperative discourse instrument. The researcher utilized an instrument based on the works of Webb and her colleagues to analyze cooperative problem-solving discourse. These are the elements of the instruments:

- Specifications for classifying help-*giving* strategies
- Specifications for classifying help-*seeking* strategies
- Specifications for debilitating behaviors

The following is a more detailed description of the instruments.

Significant research conducted over the last three decades by Noreen Webb and her colleagues suggests that cooperative learning discourse has the greatest impact on performance outcomes when learners give detailed conceptual explanations (Webb, Welner, & Zuninga, 2001), make specific and persistent requests for help (Webb & Mastergeorge, 2003a), restrain from debilitating behaviors (Webb, 1991), and apply explanations to novel problems unassisted (Webb, Troper, & Fall, 1995).

The responsiveness of peers to other peers in need, whether signaled overtly or subtly, can vary across two dimensions (Webb, Welner, & Zuninga, 2001). In one continuum is the level of help the peer tries to administer, from a good faith attempt to

deliver an in-depth, helpful answer to no response at all to an insulting or mocking response. In the other continuum is the accuracy of the response, from correct and deep to correct but shallow—all the way down to giving a wholly inaccurate response in great, misguided depth.

To capture both these dimensions, Webb, Welner, and Zuninga (2001) measured peer responsiveness on two continua: a content continuum and an intended helpfulness continuum.

Table 3. Help-Giving Continua

Level	Continuum
	Content continuum
+3	Correct answer, complete with explanation
+2	Correct answer that completely answers the question, but with little or no explanation
+1	Partially correct answer
0	No content
-1	Probably incorrect, very unclear
-2	Incorrect answer with little explanation
-3	Incorrect answer with explanation
	Intended helpfulness continuum
+3	Correct or incorrect response with explanation
+2	Correct or incorrect response with minimal explanation
+1	Correct or incorrect response with no explanation
0	Neutral
-1	Providing a response at a lower level than the capacity of the help-giver
-2	Ignoring or dismissing need for help
-3	Insulting

Note. From Webb, Welner, & Zuninga, 2001.

Webb and her colleagues never articulated the bright line between *explanation* and *minimal explanation*. From an elaboration perspective, the value of explanations is that

they force evaluation of one's cognitive structures (O'Donnell & O'Kelly, 1994). Therefore, for the purposes of this research, *explanation* was delineated as follows.

Table 4. Definition of *Explanation*

Term	Definition
Explanation	A verbal description that articulates the major principles necessary to solve a problem.
Minimal explanation	A description of only some of the major principles necessary to solve a problem, or a reference to a resource that will point a peer in a direction but does not itself describe the underlying principles or conceptual structure.
No explanation	A response that answers the question without elaborating on a reason why the answer is correct.

In terms of measuring constructive help-seeking strategies, Webb and Mastergeorge (2003a) coded behaviors into the following categories:

- General requests for help or general statements of confusion (e.g., "I don't know where to start.")
- Requests for specific explanations
- Requests for specific information ("Which question are we supposed to do next?")
- Unknowingly or questioningly articulating errors ("I got 4 as the answer [where 4 is incorrect].")

Webb and Mastergeorge's results indicated that only requests for specific explanations correlate with higher performance outcomes. The help-seeking data in the present study was coded against Webb and Mastergeorge's categories to ascertain whether their conclusions hold under the current research conditions.

In terms of debilitating group behaviors, based on Webb, Welner, & Zuninga (2001), these behaviors were flagged in this study:

- Ordering others to do something or stop doing something

- Commandeering processes (“Here, I’ll just do it.”)
- Giving insults
- Interrupting
- Off-task conversation or behavior

Another indicator of positive group functioning is unassisted problem solving after receiving help (Webb, Troper, & Fall, 1995). As much of Webb’s research focuses on the completion of sets of math or science problem, she and her colleagues were able to observe whether learners applied explanations to subsequent problems independently. Shifting the focus to larger, less structured problems makes this measurement problematic; the timeframe between receiving an explanation and the opportunity to apply it independently in the next problem is much wider. While it would be theoretically possible to track these applications, the sampling strategy in the current study was not compatible with collecting this data systematically. Therefore, this aspect of group functioning was not in scope for this study.

Semi-structured participant interviews. To gain insight into how participants viewed experience grouping, short semi-structured interviews were conducted with 10 participants. Semi-structured interviews encourage interviewees to share their insights by feeling more like a conversation than a structured interview (Patton, 2001). See Appendix A for the semi-structured interview framework used in this study.

Formal course evaluation. Every course offered by this audit firm is followed up by a voluntary, anonymous course evaluation survey for participants. These are generally distributed electronically via a trigger in the firm’s learning management system, although sometimes they are distributed on paper at the end of the course if the course is of particular interest to the firm or the target population is considered to be less likely to take the time to submit a voluntary electronic evaluation, such as executives. Internal data shows that the submission percentage for electronic evaluations during 2007 and 2008 was around 50 percent.

For comparison’s sake, the formal evaluation utilized in this study matched the evaluations typically delivered by the firm with the inclusion of a few additional questions. There were also some deviations from standard practice in order to capture

data important to the study. First, whereas evaluations for this course are usually delivered electronically, for this study the highest possible response rate was desirable so evaluations were delivered on paper. The use of paper-based surveys also allowed the design team to add additional items focused on participants' small group experiences to the stock evaluation.

An additional deviation from the stock evaluation protocols was that participants were asked to write their employee ID number on their evaluations. This deviates from the usual anonymity associated with the evaluations but was done in order to correlate satisfaction to experience group type.

While the surveys were given with all the firm's usual prompts intact, the prompt that was the focus for this research was the one considered to be the key metric by the firm in terms of measuring participant satisfaction: "Overall, I was satisfied with this course." Participants can rate their agreement to this prompt on a five-point scale from "Strongly disagree" to "Neutral" to "Strongly agree."

The following, cooperative-learning specific prompts were added to the survey by the researcher in order to tease out satisfaction with small group cooperative learning from the more holistic satisfaction prompt already present on the evaluation:

- The small group activities were effective in helping me learn.
- When I struggled, I felt comfortable asking my group for help.
- I felt like my group supported me when I struggled.
- I felt like I did a good job of helping others when they struggled.
- I asked questions in my small group.
- When I asked questions, I got detailed answers.
- When others asked me questions, I gave detailed answers.
- Overall, the small group activities were rewarding.

Participant agreement with these prompts could be indicated using the same five-point scale as the native prompts on the survey. In addition, participants were asked, "This class naturally features a wide range of incoming experience levels, from no experience to multiple audit engagements. Given your choice, what kind of small group would you prefer to be in?" Participants could choose from one of three categories:

- One made up of participants with an experience level similar to mine.
- One that includes participants with a wide range of experience.
- No opinion.

Finally, participants were given an open-ended prompt asking them about "Any other feedback you have about the small group activities."

A blank course evaluation is reprinted in Appendix B.

Experimental Design and Treatments

This study utilized a single factor experimental design where the experimental factor was the type of experience grouping utilized in forming the small groups. The levels of the factor were:

- Homogeneous, where everyone in the small group is either experienced or everyone is inexperienced, or
- Heterogeneous, where groups are a mix of experienced and inexperienced learners.

This study featured a between-subjects design.

Dependent measures. The quantitative dependent measures in this study were learning gains and satisfaction. Learning gains were measured by calculating the difference between pre-test and post-test scores. Post-test scores were not compared directly due to an expectation that more experienced learners would have a natural advantage.

Satisfaction was measured by a standard evaluation survey administered upon completion of the course. Reliability of the survey instrument was assessed using Cronbach's α.

Data analysis. Analysis of the learning gains data collected during this study consisted of two phases: an analysis of the descriptive statistics and an independent *t*-test of significance. That data was then re-analyzed—this time focused not on binary homogeneous/heterogeneous categorical predictors, but instead on experience spread across groups. For this re-analysis, a partial Pearson correlation was utilized to measure significance. The quantitative satisfaction data was analyzed utilizing a Kolmogorov-Smirnov Z Test.

The qualitative observation data was then coded utilizing the cooperative discourse instrument described above. An analysis of the descriptive statistics of the small group discourse was followed by quantitative comparisons utilizing *t*-tests for binary comparisons and Pearson correlations for continuous comparisons.

Procedures

Participants were randomized into one of two classes upon registering for the course. When participants arrived for the course, they were asked to fill out a survey summarizing their audit experience. Based on these surveys, learners were categorized as either *experienced* or *inexperienced*, with experienced auditors being defined as those having served on at least one audit engagement. Learners were then randomly placed into groups of three such that approximately half the groups would be homogeneous (all the members either experienced or all the members inexperienced) and the other half heterogeneous (one experienced auditor and two inexperienced or two experienced and one inexperienced). One group had four learners since the number of learners in that class did not divide evenly into three.

Two sections of the course were run in parallel. Each class was co-taught by one member of the firm's dedicated faculty paired with one volunteer auditor. The dedicated faculty at this firm consisted of a small number of former auditors who left fieldwork in order to teach and develop full time. One of the instructors that taught these courses had one year of experience as dedicated faculty for the organization and had previous experience teaching accounting courses at the university level. The other dedicated faculty member who taught this course was a newcomer to the role although he had taught on a volunteer basis in previous years. While the firm tries to have dedicated faculty present in as many courses as possible, the number of courses run each year in the firm far exceeds the capacity of the dedicated faculty.

The pre-test was administered at the beginning of day one and the post-test at the end of day five.

Qualitative data was collected by the lead researcher making his way systematically from group to group, observing and taking notes. Each time the activity changed, the researcher changed groups. In addition to observation, the researcher

utilized two digital voice recorders, each one situated between two members of the small group being observed, to capture the dialog for later analysis.

Late on days four and five, the research administered a series of interviews. The interviewees were chosen by purposeful sampling; during observations the researcher noted ten individuals who appeared to have experiences that exemplified some aspect of the potential pros and cons of experience grouping. For instance, one learner was chosen based on her experiences being a novice in a group with two relatively experienced auditors. Likewise, one of the experienced auditors was chosen to provide the opposite view. Of particular interest were three individuals who had already taken *Starting Your Career* at least once. The researcher sought out these individuals to provide insider contrasts between the old design of the course and the new design. Because of the use of purposeful sampling, the interview data was not meant to be representative, but instead to suggest new avenues of exploration. Each interview lasted for approximately 15 minutes and was recorded with a digital voice recorder and later transcribed by the researcher for analysis.

At the end of day five, the paper-based evaluation survey was handed out and collected in conjunction with the post-test.

Upon conclusion of the course, quantitative and qualitative data analysis began. Quantitative analyses were run using SPSS, some on a machine running version 14 and some on a machine running version 15. Learning gains and satisfaction were compared between heterogeneous and homogeneous groups. In addition, an alternative analysis was undertaken to determine if a different picture emerged when homogeneous and heterogeneous groups were redefined as ends of a continuum rather than as a binary construct.

For the observation data, the researcher utilized the coding system described above based on Noreen Webb's work. A subject matter expert, who was also an instructor and co-developer of the course, coded one of the transcripts in parallel.

During coding, all activities were first sub-divided into *exchanges*, where a single exchange was defined as a discussion on a single topic. Exchanges were initiated in one of two ways. Every time a learner asked a new, as opposed to follow up or clarifying,

question of the group, a new exchange was initiated. Or, when learners spontaneously offered information or analysis to the group without first being prompted by a question, this was also the initiation of a new exchange. The first item coded for each exchange was the nature of the initiation of the exchange.

The qualitative data was then analyzed in SPSS to compare problem-solving discourse between homogeneous and heterogeneous groupings.

The variability of problem-solving structure in the activities in this course was acute. At the more well-constrained end were problems that asked learners to answer a series of questions about the firm's core software program by exploring the program and utilizing its documentation. The questions asked in these activities typically had single, well-defined correct answers. Completing these activities generally took groups 10 to 15 minutes. At the other end of the spectrum were problem-solving activities that asked participants to generalize auditing strategies for one account to another account, a process that can take multiple hours.

In light of Merrill and Gilbert's observation (2008) that the structure and sequencing of problem solving sparks productive dialog, each activity was classified as either well-structured or ill-structured in order to determine if a decrease in problem-solving structure resulted in greater use of explanation-focused strategies. Those activities that asked learners to answer factual questions with single right answers were classified into the well-structured category and those that asked learners to engage in more open-ended problem-solving processes or to reconstruct the underlying content model in order to determine why certain problem-solving steps were valuable were classified as ill-structured. Ten of the 18 activities in the course fell into the ill-structured category.

Finally, the interviews were analyzed using a constant comparative method (Patton, 2001). The researcher read through the transcripts in multiple passes. With each pass, he would identify potential themes and then seek out supporting evidence from the responses of other interviewees.

Chapter 4: Results

This study was designed to determine the effects of experience grouping on course effectiveness. Specifically, this study was designed to answer the following research question: When professional, adult learners in problem-centered technical training courses are grouped heterogeneously and homogeneously by content experience, what is the effect on course effectiveness as measured by:

1. Learning gains
2. Problem-solving discourse
3. Learner satisfaction

Several data analyses were conducted to answer the research questions. First, the learning gains of learners in homogeneous groups were compared to the learning gains of learners in heterogeneous groups. Second, satisfaction scores were compared between the two groups. Third, both learning gains and satisfaction scores were re-analyzed to look for correlations between those variables and the spread of experience in each small group. Fourth, qualitative observation data of problem-solving discourse was coded, analyzed, and compared between homogeneous and heterogeneous groups. Fifth, participant interviews were analyzed.

Descriptive Statistics

Learning gains. Mean scores on the post-test (M = 9.76) were significantly higher than on the pre-test (M = 5.53), as determined by a dependent means t-test ($t(45) = -16.671$, $p < .001$, $r = .93$). The maximum possible score on the test was 14 points. Learning gains ranged from a low of 0.93 to a maximum gain of 9.33.

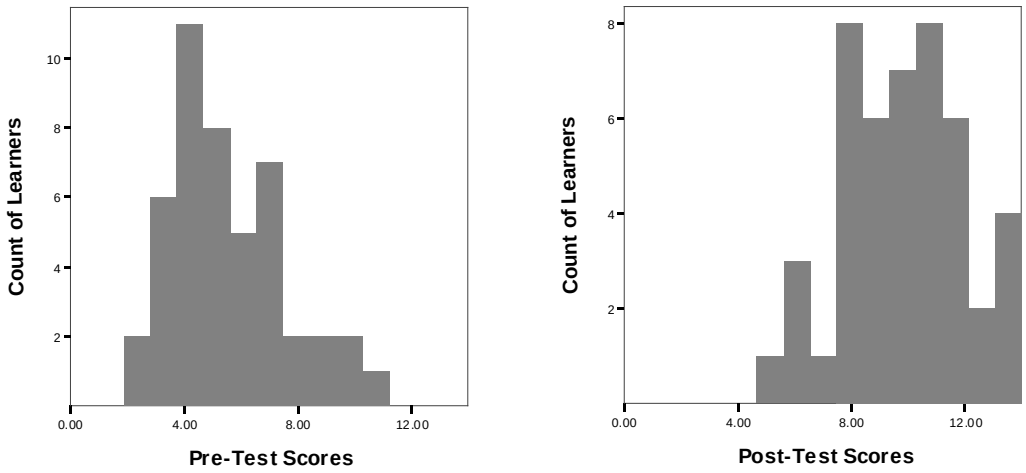

Figure 2. Comparison of Pre- and Post-Test Scores

There were no significant differences between the scores in classroom 1 and classroom 2 on the pre-test ($t(44) = .386$, $p = .701$) nor on the post-test ($t(44) = -.969$, $p = .338$).

Table 5. Scores by Classroom

	Classroom 1 Scores[a]		**Classroom 2 Scores**[b]		**Both Classes Combined**	
	M	***SD***	***M***	***SD***	***M***	***SD***
Pre-Test	5.64	2.19	5.41	1.93	5.53	2.05
Post-Test	9.48	2.32	10.07	1.72	9.76	2.05
Gains[c]	3.83	1.78	4.66	1.59	4.23	1.72

[a] $N = 24$ learners. [b] $N = 22$ learners. [c] Learning gains are calculated as post-test scores minus pre-test scores.

Experienced learners, those who had served on one or more audit engagement teams prior to the course, outscored inexperienced learners on both the pre-test and the post-test.

Table 6. Comparison by Experience Level

Test	Experience	*N*	*M*	*SD*
Pre	Inexperienced[a]	27	4.78	1.46
	Experienced[b]	19	6.59	2.31
Post	Inexperienced	27	9.45	1.80
	Experienced	19	10.18	2.34

[a] $N = 27$ learners. [b] $N = 19$ learners.

The difference in scores on the pre-test is significant ($t(28) = 3.01$, $p = .006$), even if equal variances are not assumed, as is appropriate in this case as Levene's Test for Equality of Variances is statistically significant ($F(44) = 6.27$, $p = .016$). However, scores on the post-test are not significantly different ($t(44) = 1.198$, $p = .237$), suggesting that inexperienced learners gained ground on their experienced counterparts during the course.

Learners in heterogeneous groups started with higher pre-test scores and slightly widened the gap on the post-test.

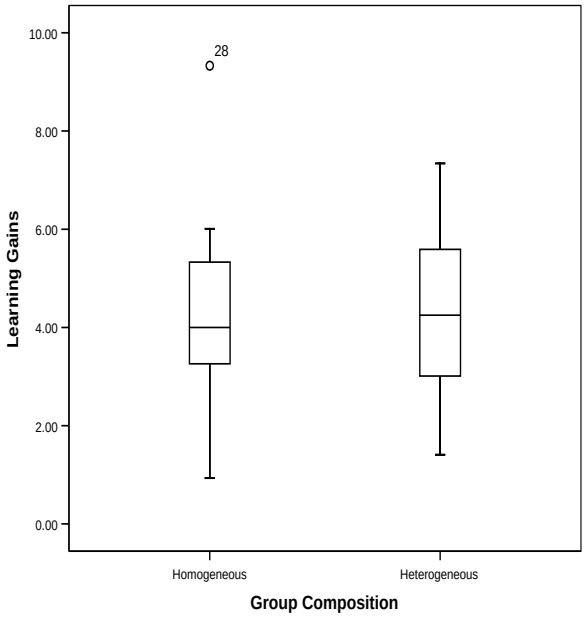

Figure 3. Learning Gains by Group Composition: Box Plot

The homogeneous group had a possible outlier as one learner had a gain of 9.33, advancing from the second lowest pre-test score in the study of 2.50 to the sixth highest post-test score, 11.83. The potential outlier's learning gain of 9.33 is almost two full points higher than the next highest gain.

Without the outlier, the mean learning gains of the heterogeneous group decreased, widening the delta between the homogeneous and heterogeneous groups.

Table 7. Learning Gains by Group Composition

Test	Group Composition	*M*	*SD*
Pre	Homogeneous[a]	5.19	2.27
	Homogeneous (sans outlier)[b]	5.30	2.25
	Heterogeneous[c]	5.94	1.73
Post	Homogeneous	9.35	2.27
	Homogeneous (sans outlier)	9.24	2.26
	Heterogeneous	10.25	1.68
Gains	Homogeneous	4.16	1.79
	Heterogeneous	4.31	1.68

[a] $N = 25$ learners. [b] $N = 24$ learners. [c] $N = 21$ learners.

None of the available data suggests a reason for the potential outlier's high learning gain. Without the outlier, the pre-test scores are not significantly different between the two groups, $t(43) = -1.05$, $p = .301$, $r = .16$. Without the outlier, the difference in post-test scores begins to approach but does not achieve significance, $t(43) = -1.66$, $p = .103$, $r = .25$. Including the outlier, differences in pre-test scores, $t(44) = -1.23$, $p = .224$, $r = .18$ and post-test scores, $t(44) = -1.50$, $p = .224$, $r = .22$ are also not significant.

The mean number of audit engagements completed prior to the course by learners in homogeneous groups ($M = 2.00$, $SD = 3.786$) was not significantly different than the mean number of engagements completed by learners in homogeneous groups ($M = 1.71$, $SD = 2.473$), $t(44) = .297$, $p = .768$, $r = .04$.

Course satisfaction. In terms of satisfaction, learners overall rated the course at $M = 4.64$ in response to the prompt "Overall, I was satisfied with this course," where a score of five indicates "strongly agree." The mean satisfaction score across all courses completed at this firm over that training year was 4.21.

Table 8. Participant Satisfaction Responses Sorted by Classroom

	Mean Participant Evaluation Scores		
Prompt	**Class 1**	**Class 2**	**Overall**
Overall, I was satisfied with this course.	4.50	4.82	4.64

The mean satisfaction score of 4.64 was significantly higher than the mean score of 4.11 that the course maintained prior to the redesign ($t(44) = 6.78$, $p < .000$, $r = .71$). Course satisfaction was also significantly higher than the firm's mean satisfaction across all courses of 4.21 ($t(44) = 5.51$, $p < .000$, $r = .64$).

Scores for both heterogeneous and homogeneous groups were also similar.

Table 9. Participant Evaluation Responses Sorted by Group Composition

	Mean Participant Evaluation Scores	
Prompt	**Homogeneous**	**Heterogeneous**
Overall, I was satisfied with this course.	4.60	4.74

Cronbach's α for the evaluation survey was calculated at .802.

Small group satisfaction. The course evaluation responses described above come from the standard course evaluations utilized by the audit firm for all courses. In addition to the standard questions, several questions specific to group work were added to the evaluation survey for the purposes of this study.

Table 10. Participant Evaluation Responses Related to Small Group Functioning by Group Composition and Classroom

Prompt	Mean Participant Evaluation Scores				
	Homo-geneous[a]	Hetero-geneous[b]	Class 1[c]	Class 2[d]	Overall
The small group activities were effective in helping me learn.	4.43	4.33	4.32	4.47	4.40
When I struggled, I felt comfortable asking my group for help.	4.43	4.61	4.45	4.58	4.47
I felt like my group supported me when I struggled.	4.48	4.44	4.50	4.42	4.43
I felt like I did a good job of helping others when they struggled.	4.35	4.33	4.27	4.42	4.35
I asked questions in my small group.	4.52	4.39	4.50	4.42	4.47
When I asked questions, I got detailed answers.	4.22	4.17	4.05	4.37	4.21
When others asked me questions, I gave detailed answers.	4.30	4.17	4.00	4.53	4.26
Overall, the small group activities were rewarding.	4.57	4.50	4.45	4.63	4.53

[a] $N = 23$, [b] $N = 19$, [c] $N = 22$, [d] $N = 20$

It appears that learners leaned toward strongly agreeing with the notion that the small group activities were valuable. While this positive perception is important, it should be tempered against findings that groups tend to overestimate their own effectiveness (Brown & Paulus, 2002).

Participants were also asked, “This class naturally features a wide range of incoming experience levels, from no experience to multiple audit engagements. Given your choice, what kind of small group would you prefer to be in?” A majority of learners chose heterogeneous groups.

Table 11. Percentage Breakdown of Responses to the Question, "Given your choice, what kind of small group would you prefer to be in?"

	N	**Percentage of Respondents**
One made up of participants with an experience level similar to mine	17	40.5%
One that includes participants with a wide range of experience	24	57.1%
No opinion	1	2.4%

The following table breaks out the data various ways.

Table 12. Category Breakdowns for Preference Data

	One made up of participants with an experience level similar to mine		One that includes participants with a wide range of experience	
	N	**Percent**	***N***	**Percent**
Class 1	7	31%	15	68%
Class 2	10	53%	9	47%
Homogeneous group members	10	43%	13	57%
Heterogeneous group members	7	39%	11	61%
Experienced	8	53%	7	47%
Inexperienced	9	34%	17	65%

Participants in class 1 expressed a preference for heterogeneous groups by more than a two to one margin, but participants in class 2 expressed a narrow preference for homogeneous groups. Overall, heterogeneous group members would prefer to stay that way, but a majority of homogeneous group members would rather switch. A narrow majority of experienced auditors would rather be grouped with other experienced auditors. Inexperienced learners, on the other hand, expressed a nearly two-to-one preference for heterogeneous groups.

Tests of Significance

This section recounts the results of the tests of significance that were run on the data summarized above in order to answer the research questions.

Does experience grouping impact learning gains? Answering the first research question requires comparing the learning gains of heterogeneous and homogeneous group members. As noted above, members of heterogeneous learning groups had slightly higher learning gains than members of homogeneous learning groups. This was unexpected; based on the available evidence that members of learning groups featuring homogeneous

ability grouping at the college level and below tend to achieve more than members of heterogeneous groups (Lou et al., 1996), the presumption in this research was that adult professionals grouped homogeneously by experience would follow the same pattern.

Also as noted above, there was one outlier in the homogeneous group. Including the outlier in the dataset, an independent *t*-test was used to compare learning gains. A check of Levene's Test for Equality of Variance was utilized to check the assumption of homogeneity of variance (Field, 2005) but the value was not significant, indicating that equal variances can be assumed. With the outlier in place, the difference in learning gains between homogeneous and heterogeneous groups was not significant ($t(44) = -.297$, $p = .768$, $r = .04$).

Of the three main options for dealing with outliers—removing the case, transforming the data, and changing the score—the best approach for reducing the impact of extreme scores is transforming the data (Field, 2005). In this case, a square root transformation of all the data produced a difference that was also not significant ($t(44) = -.363$, $p = .718$, $r = .05$).

Given these results, it cannot be reasonably concluded that grouping adults in a technical training course homogeneously or heterogeneously will impact learning gains.

Does experience grouping impact satisfaction? The second research question centered around the question of whether participating in one type of experience group leads to higher satisfaction. The key metric used by this firm to compare courses is responses to the prompt "Overall, I was satisfied with this course." Therefore, that metric was used to compare experience group types. As noted above, homogeneous groups rated the course with a mean score of 4.60 and heterogeneous groups rated the course with a mean score of 4.70 on a five point scale.

The survey data for this measure is non-parametric as all data points are integer responses, and, further, nearly all of the responses were either four or five. Since both groups had 25 or fewer members, the Kolmogorov-Smirnov Z Test was selected to compare the means (Field, 2005). Using this test, satisfaction scores between homogeneous and heterogeneous groups were not significantly different ($Z = .200$, $p = 1.00$, $r = .03$). Therefore, it cannot be reasonably concluded that being in a heterogeneous

or homogeneous group based on audit engagement experience is more likely to lead to higher satisfaction in a technical training course for professional adults.

Tests of Significance, An Alternative Approach

The analysis thus far has been predicated on the assumption that experience is a binary construct—either someone is experienced or they are not. An alternative way to look at experience grouping is from a continuum perspective rather than a binary perspective. In other words, the dynamics in a small heterogeneous group where the most experienced member has been on only a single engagement may be much different than if that person has been a member of fifteen audit engagements. Or, a group with two individuals with no experience and one with one engagement may perhaps feel much more homogeneous than a group with one learner with three engagements, one with four, and one with twelve.

To test this alternate perspective, the tests of significance were re-run, but instead of categorizing groups on a binary scale, the groups were recharacterized along a continuum based on the spread of experience in the group. The spread was calculated by subtracting the number of audit engagements of the least experienced group member from the number of audit engagements of the most experienced member. A group with a learner with no audit engagements, one with three engagements, and one with 15 engagements would therefore have a spread of 15 engagements. A group with three learners all with four audit engagements would have a spread of zero.

Recharacterized this way, groups ranged from a spread of zero engagements to a maximum of 12 engagements.

Table 13. Number of Small Groups at Each Level of Experience Spread

Experience Spread	Number of Groups
12	1
11	
10	
9	
8	1
7	
6	2
5	
4	2
3	
2	2
1	2
0	5

A scatterplot suggests a negative relationship between experience spread of a learner's group and that subject's learning gains.

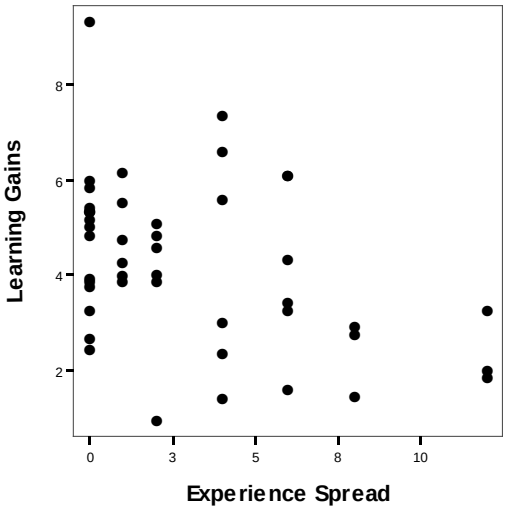

Figure 4. Relationship Between Learning Gains and Experience Spread

The group experience spread data is skewed right and truncated at zero and is therefore non-parametric. Therefore, an appropriate test of significance is Spearman's correlation (Field, 2005).

Spearman's correlation indicates there is a statistically significant negative moderate correlation between individual learning gains and group experience spread ($r_s(46) = -.369$, $p = .012$). As the experience spread inside learning groups increases, learning gains decrease. The correlation is identical when the data is logarithmically transformed to compensate for the potential outlier.

One potential confound in this alternative analysis is that the high end of the experienced spread scale has a greater percentage of experienced learners. This is a potential concern due to a possibility of a ceiling effect—one would expect more experienced learners to have on average higher pre-test scores, giving them less room to grow than inexperienced learners. This might weigh down average scores in groups with large experience spreads. Indeed, pre-test scores have a significant positive relationship with experience spread ($r_s(46) = .594$, $p < .001$).

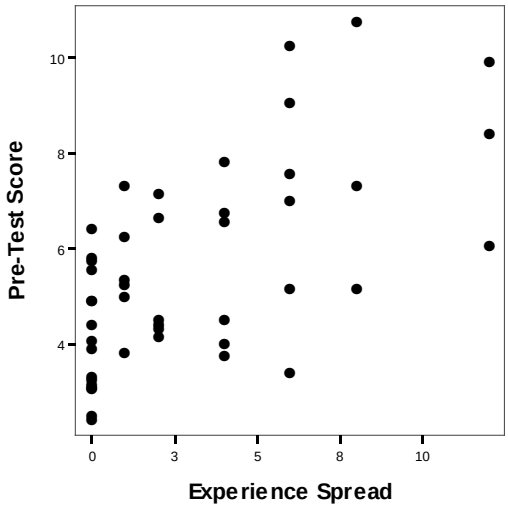

Figure 5. Relationship Between Experience Spread and Pre-Test Scores

One approach to dealing with the potential confound is to run a partial correlation comparing experience spread and post-test scores while controlling for pre-test scores. Because SPSS v14.0 does not offer partial Spearman correlations, the three variables were rank ordered and then a partial Pearson correlation was generated. Correlating post-test scores with experience spread while controlling for pre-test scores, the correlation drops below the level of significance ($r_s(43) = -.206$, $p = .175$).

Given these analyses, one cannot conclude that experience spread in a small learning group composed of adults in a technical training course impacts learning gains.

Qualitative Analyses

Two levels of qualitative data were collected. First, the researcher observed and recorded the audio of different groups during each small group activity during the course. Second, ten participants participated in semi-structured interviews.

Help-seeking and debilitating behavior codes. Webb and Mastergeorge (2003a) coded four types of "indication of need for help" in their study of small group learning around a math problem asking learners to calculate the cost of a telephone call when the first minute is charged at a different rate than subsequent minutes, as described in the table below.

Table 14. Help-Seeking Behaviors as Defined by Webb and Mastergeorge

Indication of Need for Help	Definition According to Webb and Mastergeorge
General request for help	"Indications of a need for help included general requests for help about how to solve the problem ('How did you get that one?') or general statements of confusion ('I don't get it')" (p. 372)
Request for specific explanations	"Requests for specific explanations about obtaining the number of additional minutes ('How did you get 29?')" (p. 372)
Request for information	"Requests for specific information concerning the number of additional minutes ('Are you sure it's 13 times 29?')" (p. 372)
Making an error	A learner makes an error verbally.

Note. Based on information presented in Webb and Mastergeorge (2003a)

Webb and Mastergeorge, by virtue of studying group dynamics around a single, well-focused class of problems, were able to define a division between asking for specific explanations and merely requesting information relative to that type of problem. The present study analyzes small group responses across a number of disparate problems and therefore requires a more generalizable framework. Therefore, for purposes of this study, Webb and Mastergeorge's four categories were generalized as shown in the table below. In addition to their four categories, two additional categories were added to capture situations not covered by their framework.

Table 15. Help-Seeking Codes Used in this Study

Indication of Need for Help	Definition for this Study
General request for help	A verbal declaration of a need for help that offers no indication of the nature of the learner's confusion. For example, declaring "I'm

Indication of Need for Help	Definition for this Study
	so confused!" tells potential help-givers that the learner needs help, but offers no indication of the type of help needed.
Request for specific explanations	A verbal request for help that indicates a desire to understand the underlying conceptual model—in other words, the help requestor appears to want to the help giver to give not just an answer but to explain *why* it works that way. This is different than merely seeking a yes/no answer or a confirmation or merely asking for a series of steps. Examples from the data collected include: • "How come you're doing 105?" The learner isn't asking how to do it, but why it needs to be done. • "Dude, I'm curious **why** I can't see that revenue on my trial balance. 20 million. I'm wondering if we imported the right trial balance or whatever." If he had just asked, "Did we import the right trial balance?" then this would be coded as seeking information rather than an explanation. • "Could you help me? I'm really not understanding that importing deal." This is close to a general request for help but is differentiated by the fact that he isolates the specific source of trouble. If he had instead implored, "Could you help me? I'm really not understanding any of this," then it would have been a general request for help. Likewise, since he notes that he is seeking understanding rather than just the right answer or the steps, this implies that he is requesting an explanation rather than just information. • In response to an apparent discrepancy in the data, a learner comments: "Geez, these guys are way too efficient. They bill

Indication of Need for Help	Definition for this Study
	before they ship." This declaration implies a wish to understand how that's possible. • "It says we have to import the trial balance. But it's already here." Again, the learner appears puzzled and to wish to understand the contradiction.
Request for information	A verbal request for help that does not necessarily imply a desire to receive an explanation of *why* the system works that way. Yes/no questions typically fall into this category, as would questions asking for the steps of how to complete a task. Examples: • "You said you went into the 200 file to work in the financial statements?" If he had asked, "Why did you go into the 200 file?" then it would have been coded instead as a request for a specific explanation. The question as stated, however, invites merely a yes/no response. • "Trace from the outstanding check list to the bank balance, or the bank balance to the outstanding check list?" This question invites a short factual answer rather than an explanation. • "I've got a question for you, would we need to like for our outstanding checks for the operating accounts, would we need to bring this in to like an actual document, or can I use this like in the paper file or whatever?" The question invites the help-giver to answer either/or, rather than *why* she would use one over the other. The help-giver may choose to explain why, but could answer the question without giving an explanation. Note that requests for information do not have to be direct

Indication of Need for Help	Definition for this Study
	questions. If a learner offers an accurate description or explanation to one or more members of the group unprompted with an implied "is that correct?" then it would be coded in this category. A similar offering that is *incorrect* would be coded in the *making an error* category, below.
Making an error	A verbal offering of information or an explanation that another learner corrects, even if the help-giver him or herself is in error. Example: • "Back of files…I know that one. 300 to [garbled]," to which a help-giver made a wrong answer buzzer noise, setting the stage for an explanation of why the answer is wrong. • "Hey, we got something different." Here, the help-seeker discovered a discrepancy, inviting a discussion to achieve resolution.
Implied request for help	All of the code above assume that a help-seeker is initiating the exchange. An exchange may be initiated by a help-giver who senses that one or more learners in the group needs help, even if the one who needs help does not even know it yet. This sense might come from inaudible signals, from observing their work, from knowledge that comes from working with the other learners in the group, or even from prior knowledge about the topic that would cause the help giver to believe that others will need help. Therefore, exchanges that begin with a description or explanation rather than a question could fall into one of three categories: (1) a *request for information* if the individual initiating the exchange

Indication of Need for Help	Definition for this Study
	offers accurate information and appears to be seeking confirmation, (2) *making an error* if the individual seeks confirmation for incorrect information, and (3) *implied request for help* if the exchange initiator appears to be offering information not to confirm its accuracy but in order to help others that he or she perceives need the information.
Related	Sometimes exchanges are initiated that are not help-seeking nor help-giving regarding the problem being solved but are still germane. For instance, a learner might ask other group members, "Should we all just go through it together?" The answer doesn't directly help anyone solve the problem at hand, but it is still on task. Likewise, if one learner asks another, "How many engagements have you been on?" the information sought isn't directly helpful to solving the immediate problem—but questions *can* be germane to the objectives of the course if they lead to conversations about what audit engagements are like and what they should expect and how they can prepare. Thus, any kind of potentially useful but not directly related to the problem at hand exchanges were coded in this category.

Of course, exchanges can also be initiated that do not directly advance learners toward fulfilling the course objectives. Webb, Welner, and Zuniga (2001) call these *debilitating behaviors.* This label implies that off-task behavior is devoid of benefits toward problem solving; Webb and her colleagues have not published research exploring the effects of debilitating behaviors in building camaraderie and the effects these have on group dynamics and problem solving over the long term. In line with Webb's research, these behaviors were coded as debilitating:

1. Ordering others to do something or stop doing something
2. Commandeering processes (e.g., “Here, I’ll just do it.”)
3. Giving insults
4. Interrupting
5. Off-task conversation or behavior

Help-giving codes. *Exchanges* are two-way processes; a help-seeker expresses, in various ways, a need for help and then help-givers respond or choose not to respond. The responses to the help-seekers were coded along two continua, again in line with Webb’s work (Webb, Welner, and Zuninga, 2001). The first continuum describes, over the course of the entire exchange, the highest level of *intended* helpfulness. As noted in the literature review, Webb and her colleagues defined this continuum along a seven point scale from +3 (correct or incorrect response with explanation) to 0 (neutral) to -3 (insulting). Examples of how the intended helpfulness continuum were applied in this research are described in the table below.

Table 16. Intended Helpfulness Codes

Score	Description
+3	**Correct or incorrect response with explanation**
	In one exchange, a learner made a *request for specific explanation*: “What are these miscellaneous items?” (A query such as “Do we do anything with these miscellaneous items?” would have been a *request for information* since it would have invited a simple negative answer or a description of a procedure.) The help-giver in the group began explaining *scope*, a concept that describes how big a transaction has to be before the auditors need to scrutinize it. His initial answer featured *minimal explanation* (+2, see below): “Miscellaneous? Again, scope. That’s 15,000; you can ask them about it.” But in response to follow-up questions, the help-giver gave more in-depth explanations such as: “It’s under our scope, it’s a single transaction and it’s within 3% of [the materiality limit],

Score	Description
	so it falls within that 3% therefore we have to go to the controller and that could be a check that is now [unclear]; it's already on the reconciliation, we just need to verify what that is. They've noted it, we just need to know what it is. If it is a check that they've held back, then that's great. You know, then we already know what it is, they've already booked it, they've brought it back into the period, they've brought it back into the bank balance. In the books. But if it's multiple transactions then we're not going to worry about it because then it's under the ten-five. Good catch." With this response, whether right or wrong, the help-giver is attempting to explain the underlying conceptual model. This elaborated response therefore was coded +3.
+2	**Correct or incorrect response with minimal explanation** In one activity, learners were asked why particular documents can't be changed. One help-seeker made a *request for information*, "It's just cause it's the year-end report, right?" to which another responded, "Yeah, it's a canned document because it's an automatic document—all automatic documents are canned." The help-giver did not explain the relationship between automatic documents and manual documents and how the information flows through them, which would have warranted an intended helpfulness score of +3, but neither did he merely answer the question in the affirmative and leave it at that, which would have been +1. His response was somewhat circular—basically saying that canned documents are canned—but he was trying at least to offer some elaboration.
+1	**Correct or incorrect response with no explanation** Part of one activity asked learners to summarize the problem-solving steps. One learner began summarizing by noting, "So, number one, what we did was take bank confirmations," which was in error. A help-giver corrected him, noting only "We looked at the leadsheet. Number one, we looked at the leadsheet." He did not attempt to explain *why* looking at the leadsheet was the appropriate first step nor why taking bank confirmations was not.

Score	Description
0	**Neutral** Sometimes, it was not clear what level of help the help-giver was intending to give. For instance, one help-seeker implored, upon receiving an unexpected response by a computer program, "What did I do?!" A potential help-giver asked an initial clarifying question, "Are you on number 7?" To which the help-seeker responded, "I tried to—I just shut [the application], dammit." The help-giver laughed and the exchange ended there as the help-seeker had to go through the process of re-starting everything. The thread was never picked up again, leaving it unclear what level of help the help-giver would have tried to provide. This category was also used when help-givers respond that they don't know the answer. They are not providing help, but neither are they ignoring the plea.
-1	**Providing a response at a lower level than the capacity of the help-giver** This particular code was rarely used. In Webb's research, this code seems intended to describe the behavior where a help-giver is holding back because he or she simply doesn't want to help more even though he or she could—for instance, to stem a perceived freeloader effect where a help-giver feels like he or she is doing all the work while the help-seeker will get just as much credit. There was little evidence of that kind of withholding of information during this course. While there were instances where help-seekers did not receive meaningful help, there was no way for the researchers to discriminate whether that lack of help was because the potential help-givers were unwilling or unable to help. Therefore, instances where help-seekers truly received no response, the *intended helpfulness* was coded as -2. The one clear instance of a help-giver granting sub-optimal help was when a help-giver deliberately told a help-seeker the wrong answer and then later admitted she was just kidding. This kind of behavior, at least when groups were being observed, was extremely rare over the course of the week-long class.
-2	**Ignoring or dismissing need for help**

Score	Description
	This category was used when a help-seeker received no response to an appeal for help.
-3	**Insulting** This category would have been used to tag any rude or denigrating behavior in response to pleas for help. However, no instances of insulting behavior were observed.

Note that only learners received intended helpfulness codes. When instructors were the help-givers, no code was given.

Intending to be helpful does not guarantee that the help-giver is actually helpful. An explanation may simply be wrong. Therefore, each exchange was also coded to reflect the actual level of help conveyed. Based on Webb, Welner, and Zuninga (2001), the table below describes how each code was applied to the data in this study.

Table 17. Level of Help Conveyed

Score	Description
+3	**Correct answer, complete with explanation** In one exchange, a help-seeker was confused about whether he was looking at this year's audit workpapers or last year's. The help-giver affirmed correctly that the workpapers were last year's and then elaborated about the importance of the difference: "Yeah, these are prior year's. We're still preparing to do a cash audit. Now, if we were actually auditing cash we'd have to have, well, we have the confirms, we'd also need to have probably bank statements, and we wouldn't have already done the work on the client [unclear] engagements."
+2	**Correct answer that completely answers the question, but with little or no explanation** In one exchange, a help-seeker declared, correctly, that he thinks they should be

Score	Description
	using a particular form. The other members of the group agreed without further elaboration. In another exchange, a help-seeker asked whether he is using the correct version of the software. The help-givers confirmed that he was not. They also give him step-by-step instructions on how to upgrade. However, the conversation stays at the procedural/factual level. The group members never attempt to explain why the system works as it does (e.g., why the old version should be removed before installing the new version).
+1	**Partially correct answer** This code indicates that the efforts of the help-givers were on the right track but were incomplete or only partly accurate.
0	**No content** This code was chosen when a help-seeker received no answer at all or no useful answer, such as "I don't know."
-1	**Probably incorrect, very unclear** On occasion, a help-giver's answers were confusing and disjointed, leaving coders unable to really follow their logic. This code was assigned in those situations.
-2	**Incorrect answer with little explanation** Similar to +2, except that the answer was conceptually incorrect.
-3	**Incorrect answer with explanation** Similar to +3 above, except that the answer was conceptually incorrect.

Exchanges where instructors were the help-givers were not coded on the above scale.

Descriptive statistics of qualitative data. The researcher developed the coding system described above and the both the researcher and the lead subject matter expert coded one of the most extensive transcripts in parallel. Then the coders compared their work and talked through discrepancies until they reached agreement on the most appropriate codes. The researcher coded the rest of the transcripts independently.

Over four days of the course—the first day, which had the fewest problem-solving activities, was not recorded successfully due to technical difficulties—472 separate exchanges containing a total of almost 5000 separate conversational turns—one person speaking until he or she stopped or was interrupted—were coded over 18 separate problem-solving activities. These figures include two activities in which two different groups were recorded and observed due to different pacing across the two classrooms. The problem-solving activities lasted from a low of 9 minutes to multiple hours—for problem-solving activities lasting longer than an hour only the first hour was recorded. The groups averaged about 54 exchanges per hour, though this varied from about 26 to about 90 exchanges per hour. Exchanges averaged about 10.5 turns per exchange.

In terms of help-seeking, by far the most common strategy was the posing of a request for information, as 56% of all exchanges were initiated in this fashion.

Table 18. Help-Seeking Observed

Level of Help-Seeking	***N***	**Percentage**
General request for help	16	4%
Request for specific explanations	75	18%
Request for information	233	56%
Making an error	7	2%
Implied request for help	50	12%
Related	37	9%

Help-seekers were three times more likely to ask for information than for a specific explanation.

The learners appeared to take the problem solving very seriously, at least while they were being recorded. Only 12 instances of debilitating behavior were recorded, and fully half of them were from one group. All of the debilitating instances were off-task conversations; there were no recorded instances of insulting or other uncivil behavior.

Informal observations of groups that were not being recorded or formally observed did not uncover any evidence that groups acted differently in terms of debilitating behaviors when they were not being observed.

In response to help-seeking, help-givers were more likely to offer a correct or incorrect response with minimal explanation than they were to offer a correct or incorrect response with an explanation. In fact, help-givers were more likely to give a correct or incorrect response with no explanation than they were to give an explanation. That said, the intended helpfulness percentages are more evenly distributed than the help-seeking percentages.

Table 19. Levels of Observed Intended Helpfulness

Level of Intended Helpfulness[a]	***N***	**Percentage**
Correct or incorrect response with explanation	71	21%
Correct or incorrect response with minimal explanation	115	34%
Correct or incorrect response with no explanation	94	28%
Neutral or lower	59	17%

[a] Instructor responses excluded.

Similarly, “correct answer that completely answers the question, but with little or no explanation” was the most common level of help ultimately received.

Table 20. Levels of Observed Help Received

Level of Help Received[a]	*N*	Percentage
Correct answer, complete with explanation	54	17%
Correct answer that completely answers the question, but with little or no explanation	145	45%
Partially correct answer	50	16%
Neutral or lower	70	22%

[a] Instructor responses excluded

Instructors acted as help-givers in 63 of the exchanges, about 13% of the total number of exchanges.

On the whole, then, learners in this study tended to request information rather than explanations, and generally that is what they received. However, those percentages are not necessarily meaningful in the abstract—just because learners asked three informational questions for every one explanation-seeking question does not necessarily mean that they did not utilize explanation-seeking questions appropriately for the problems. However, if heterogeneous or homogeneous learning groups systematically sought or received explanations more than the other type of group, that would constitute evidence, given the work of Webb and her colleagues, that that one type of group was more effective at supporting problem solving.

Comparing heterogeneous and homogeneous groups in terms of problem-solving dialog. As shown in the table below, homogeneous learning groups tended to request a slightly higher percentage of specific explanations and intended to deliver a higher percentage of explanations, though heterogeneous groups were more successful at ultimately delivering correct explanations. None of these differences were statistically significant, however.

Table 21. Percentages of Positive Problem-Solving Approaches by Group Type

Problem-Solving Approach	Homogeneous Learning Groups[a]		Heterogeneous Learning Groups[b]	
	M	***SD***	***M***	***SD***
Help-seeking = Requests for specific explanations	20.9%	14%	17.5%[c]	24%
Intended helpfulness = Correct or incorrect response with explanation	27.3%	29%	24.4%[d]	34%
Level of help received = Correct answer, complete with explanation	21.4%	22%	24.6%[e]	29%

[a] Ten homogeneous groups were observed

[b] Eight heterogeneous groups were observed

[c] Differences in help-seeking were not significant ($t(16) = .371$, $p = .716$, $r = .09$)

[d] Differences in intended helpfulness were not significant ($t(16) = .191$, $p = .851$, $r = .05$)

[e] Differences in help received were not significant ($t(16) = -.260$, $p = .798$, $r = .06$)

Another way to view the data is in terms of raw rate of explanation-seeking and explanation-giving—it may not be the percentage of exchanges result in explanations, but rather the raw number of explanations sought and delivered that matters. The table below shows a comparison of rate per hour of explanation-seeking and explanation-giving behaviors by group type.

Table 22. Rate of Problem-Solving Approaches Per Hour by Group Type

Problem-Solving Approach	Homogeneous Learning Groups		Heterogeneous Learning Groups	
	M	***SD***	***M***	***SD***
Help-seeking = Requests for specific explanations	9.47	6.60	6.85[a]	8.31
Intended helpfulness = Correct or incorrect response with explanation	9.25	6.48	8.40[b]	11.44
Level of help received = Correct answer, complete with explanation	7.28	6.04	8.25[c]	9.59

[a] Differences in help-seeking were not significant ($t(16) = .745$, $p = .467$, $r = .18$).

[b] Differences in intended helpfulness were not significant ($t(16) = .198$, $p = .845$, $r = .05$).

[c] Differences in help received were not significant ($t(16) = -.263$, $p = .796$, $r = .07$).

Raw rate of explanation-seeking and -giving behaviors tracked closely to explanation-related percentages. Homogeneous groups tended to ask for and intended to grant more explanations per hour and heterogeneous groups tended to successfully give more explanations per hour. However, the differences were not statistically significant.

Problem-solving dialog as function of experience spread. Another possibility is that any differences in explanation-focused exchanges may emerge only as a function of the spread of experience across groups—as the spread of experience in a group grows, the most experienced members may be more or less willing to provide explanations, and less experienced members may be more or less willing to ask for them. The scatterplots below show the relationships between experience spread in groups and explanation-focused exchanges.

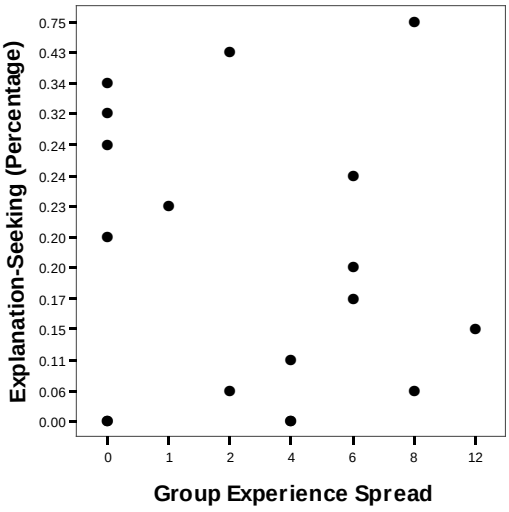

Figure 6. Percentage of Help-Seeking Attempts that Appeal for Explanations as a Function of Group Experience Spread

Note. Correlation is not significant (r = .117, p = .643)

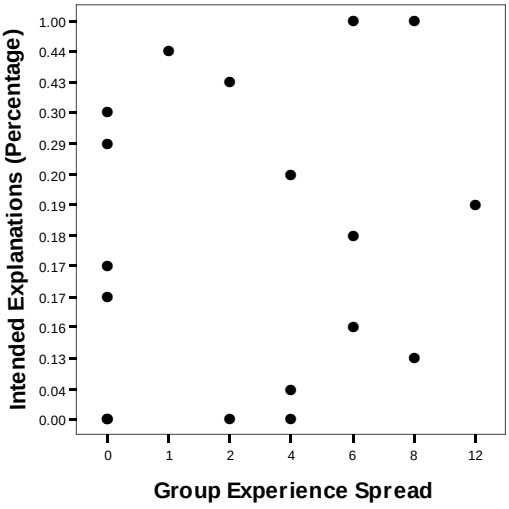

Figure 7. Percentage of Help-Giving Responses Intended as Explanations as a Function of Experience Spread

Note. Correlation is not significant (r = .280, p = .261)

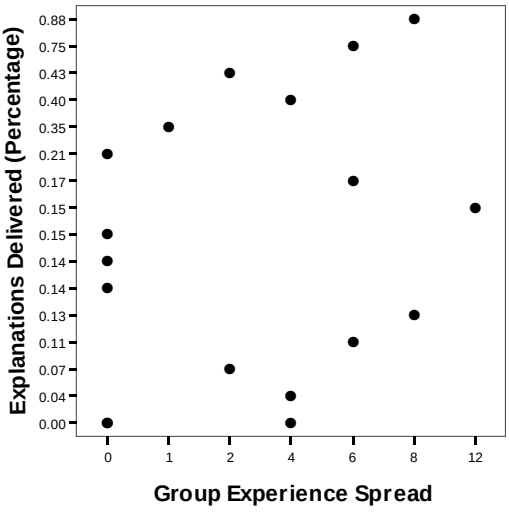

Figure 8. Percentage of Help-Giving Responses that Include Accurate Explanations as a Function of Group Experience Spread

Note. Correlation is not significant (r = .314, p = .204)

Pearson correlations reveal no statistically significant correlations among the percentages of help-giving and seeking behaviors and the experience spread in the groups.

Role of problem-solving structure. In light of Merrill and Gilbert's observations (2008) regarding the effects of problem solving on discourse, the data was split to allow comparison of ill-structured problem-solving dialog and well-structured problem-solving dialog. Ill-structured problem solving resulted in greater use of explanation-focused problem-solving strategies, as shown below.

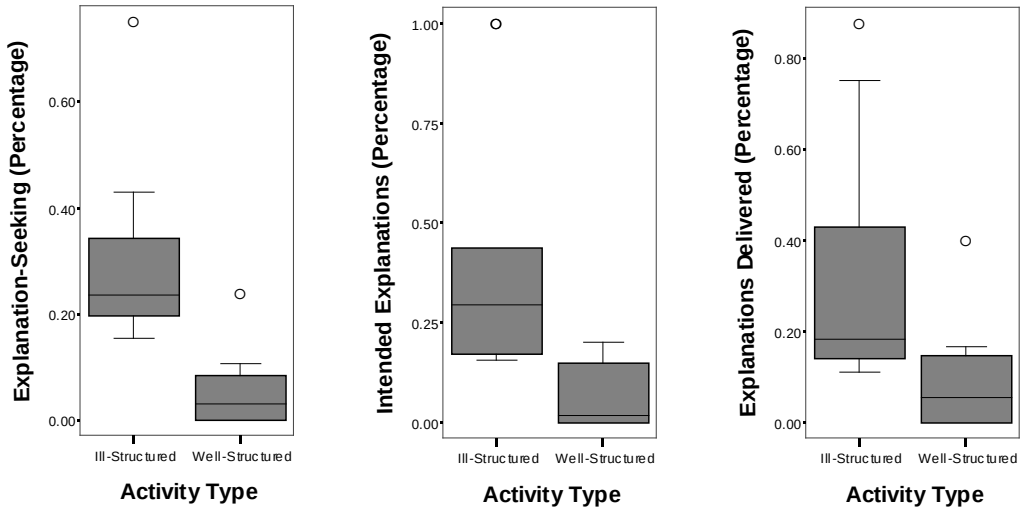

Figure 9. Explanation-Focused Strategies Utilized for Different Problem-Solving Structures

The differences between explanation-focused strategy use solving ill-structured problems as compared to solving well-structured problems were significant. Groups solving ill-structured problems sought a significantly greater percentage of specific explanations (M = 30.2%, SD = 17.9%) than groups solving well-structured problems (M = 5.9%, SD = 8.3%), $t(16) = 3.54$, $p = .003$, $r = .66$. They also intended to give a significantly higher percentage of explanations (M = 41.5%, SD = 32.5%) than groups solving well-structured problems (M = 6.7%, SD = 8.6%), $t(16) = 2.93$, $p = .010$, $r = .59$. Finally, they succeeded in delivering a significantly greater percentage of correct answers complete with explanations (M = 33.1%, SD = 27.5%) than did groups solving well-structured problems (M = 10.0%, SD = 13.6%), $t(16) = 2.16$, $p = .046$, $r = .48$.

The above analysis provides evidence in favor of Merrill and Gilbert's assertion (2008) that less-structured problem solving leads to higher level small group functioning. In terms of analyzing the differences between homogeneous and heterogeneous groups, it also suggests that ill-structured and well-structured problem solving should be analyzed separately.

Re-analyzing differences in explanation-focused strategy use by focusing on ill-structured problem solving. Limiting the dataset only to the ten ill-structured activities, heterogeneous groups did come out higher in all three explanation-focused categories. In fact, heterogeneous groups exchanged explanations at a much higher rate than the homogeneous groups—however, since the experiment was not originally designed to vary this attribute systematically, only three of the ten observed ill-structured problem-solving sessions featured heterogeneous groups. The limited dataset compromised statistical power, making it difficult to draw conclusions.

Table 23. Comparing Heterogeneous and Homogeneous Explanation Focus during Ill-Structured Problem Solving

	Homogeneous Learning Groups[a]		**Heterogeneous Learning Groups**[b]	
Explanation Focus	***M***	***SD***	***M***	***SD***
Help-seeking = Requests for specific explanations	26.5%	10.2%	39.0%[c]	31.1%
Intended helpfulness = Correct or incorrect response with explanation	36.4%	29.5%	53.2%[d]	42.9%
Level of help received = Correct answer, complete with explanation	28.2%	23.1%	44.4%[e]	22.6%

[a] $N = 7$

[b] $N = 3$

[c] Differences in help-seeking were not significant ($t(8) = -1.01$, $p = .341$, $r = .33$).

[d] Differences in intended helpfulness were not significant ($t(8) = -.727$, $p = .488$, $r = .25$).

[e] Differences in help received were not significant ($t(8) = -.893$, $p = .426$, $r = .30$).

In terms of experience spread instead of binary categories, explanation-seeking appeared to exhibit a parabolic relationship with group experience spread.

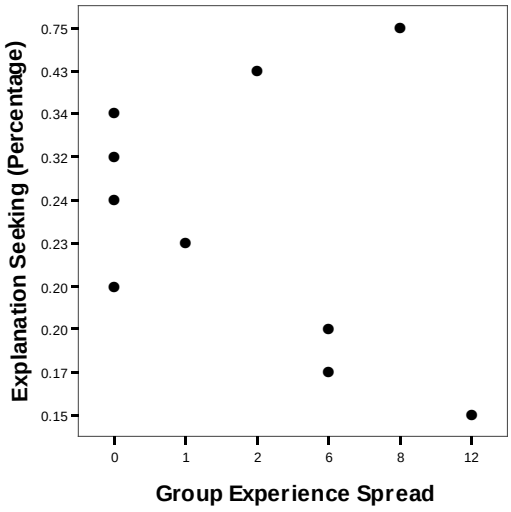

Figure 10. Percentage of Help-Seeking Events that Were Explanation-Seeking as a Function of the Spread of Experience in the Group

Fitting a quadratic curve to the data resulted in a relationship that has a large effect size and is borderline significant, $R^2 = .567$, $F(2, 7) = 4.587$, $p = .053$. This finding suggests that greater experience spreads may polarize help-seekers, either driving down their willingness to seek explanations or driving it up.

The relationship between explanation-giving and experience spread, however, while still suggesting possible parabolic relationships, are farther from statistical significance.

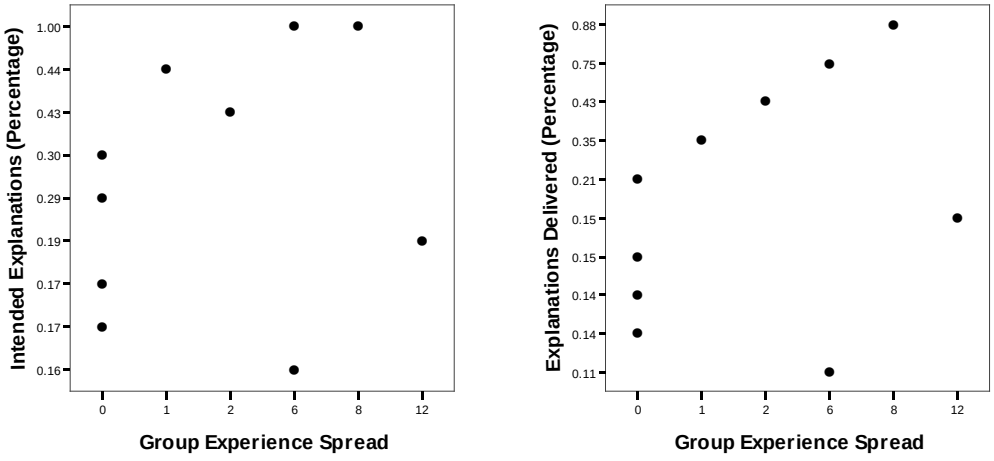

Figure 11. Explanation-Granting in Ill-Structured Problem Solving as a Function of Experience Spread in the Group

A potential quadratic relationship between intended explanations and group experience spread was not significant (R^2 = .290, $F(2, 7) = 1.433$, p = .301), neither was a possible quadratic relationship between percentage of correct explanations delivered and experience spread (R^2 = .244, $F(2, 7) = 1.130$, p = .376).

Relationship of explanation-focused exchanges to learning gains. The qualitative analysis completed for this study was focused on explanations due to the extensive work of Webb and her colleagues showing that giving explanations correlates with increased achievement (e.g., Webb, 1982b, 1991) and that asking for specific explanations also correlates with increased achievement (e.g., Webb & Mastergeorge, 2003a, 2003b). However, there is no existing evidence that those correlations hold with professional, adult learners in technical training courses. Nor is there strong evidence that the correlations hold in ill-structured problem-solving situations. Therefore, it is worth exploring whether the present study offers evidence in support of extending Webb's correlations to this study's population.

Surprisingly, explanation-giving seems to correlate negatively with learning gains—that is, small groups that both intended to give more explanations and successfully gave more explanations had lower learning gains, at least on first analysis.

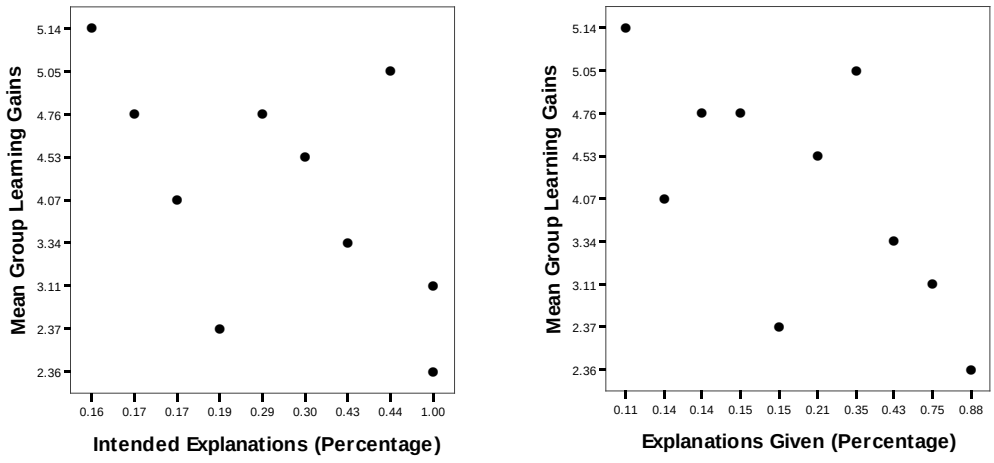

Figure 12. Relationships between Explanation-Giving and Mean Learning Gains of Small Groups

Even with the limited dataset created by restricting the analysis to the ill-structured problems, these negative correlations begin to approach statistical significance when analyzed using Pearson correlations. As the percentage of intended explanations rose, the average learning gains for the group fell (r = -.556, p = .096); likewise, as the percentage of responses that were correct answers with explanations rose, average learning gains for the group also fell (r = -.614, p = .058). The effect size in both cases was large. It is worth noting that while learning gains fell as explanations increased, it appeared there was no relationship between learning gains and the rate of exchanges—that is, it did not appear that groups who talked more or less scored higher or lower (r = -.046, p = .450). Nor did there appear to be a correlation between the number of individual offerings per exchange and learning gains (r = -.010, p = .978).

Much like the analysis earlier, however, where experience spread negatively correlated to learning gains until pre-scores were factored in, it appears that prior knowledge is a confound here as well. Re-running the analysis, this time utilizing a partial Pearson correlation to factor out pre-test scores, reduced the correlations. Partialing out the pre-test scores reduced the correlation of learning gains to both intended explanations (r = -.315, p = .410) and to correct explanations delivered (r = -

.399, p = .287) to the point where they no longer approach significance. Therefore, while there appeared to be a moderate negative correlation between explanation-giving and learning gains, the correlation were not statistically significant and no conclusions could be reached.

The picture for explanation-seeking was interesting, as it appeared as two crisscrossed lines or perhaps a parabola—though with limited data this could easily have been an illusion.

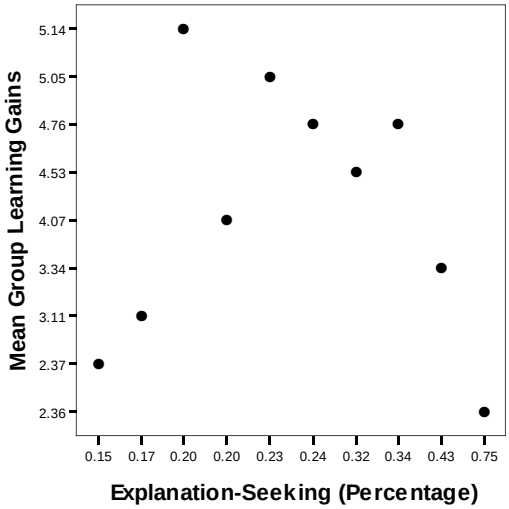

Figure 13. Relationship of Mean Learning Gains for a Group and the Percentage of Help-Seeking Events Focused on Receiving Explanations

A Pearson correlation reported a negative effect of medium size—as groups sought more explanations, their mean scores tended to drop—but the effect was not statistically significant (r = -.385, p = .272). Further, attempting to fit a quadratic curve to the data showed a strong but statistically insignificant correlation (R^2 = .417, $F(2, 7) = 2.50$, p = .152). Given the lack of significance, this study does not provide evidence that increased help-seeking correlates with learning gains.

Role of instructor explanations. Though the correlations between learning gains and explanation-seeking and receiving did not achieve levels of statistical significance, it was unexpected that the correlations would be generally negative. Should data in future

studies bear out this negative correlation, one possible explanation could be that learners who exchange more explanations with peers received fewer from instructors, a dynamic not addressed in Webb's work. In the present study, the help-giver was the instructor 16.8 percent of the time during ill-structured problem-solving activities, either because the group appealed to the instructor or, less often, because a listening instructor intervened when a group was having trouble.

Pearson's correlation showed a moderate positive correlation between percentage of teacher help-giving and learning gains. This correlation was not, however, statistically significant (r = .359, p = .154).

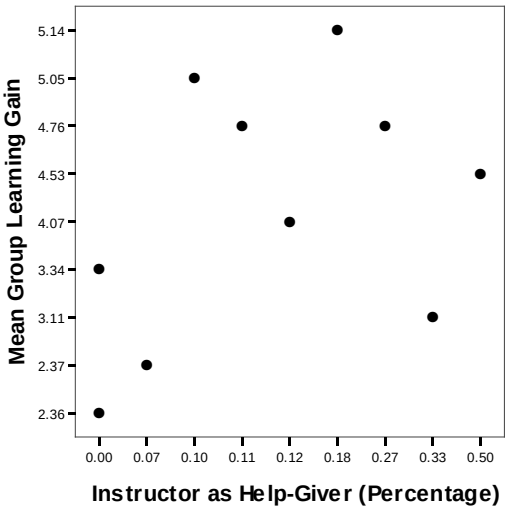

Figure 14. Mean Group Learning Gains as a Function of the Percentage of Exchanges Where the Instructors is the Help-Giver

There was also a weak negative correlation between percentage of instructor explanations and percentage of peer explanation, though this correlation was also not statistically significant (r = -.108, p = .384).

In terms of the relationship between peer explanations and instructor explanations, the data collected for this study suggests that there may be a moderate negative correlation between peer explanations and learning gains and a moderate positive correlation between instructor explanations and learning gains. Those

suggestions are speculative at this point as they do not meet applicable standards for statistical significance.

Interviews

The following themes emerged from the interview data.

Table 24. Interview Themes

Group	Coded Theme
All	Positive themes
	Appreciation for the opportunity to work in groups
	Negative themes
	Trade-offs in any group arrangement
Heterogeneous	Positive themes
	A preference for heterogeneous grouping
	An appreciation for playing the role of the tutor
	Negative themes
	Difficulty keeping up with more experienced groupmates
Homogeneous	Positive themes
	An appreciation for camaraderie of the cohort
	Negative themes
	Would rather have been in a heterogeneous grouping
Interviewees who had taken the course before	Positive themes
	Appreciation of the significant scaffolding and focus on continuous problem solving in the redesigned course
	Negative themes
	None noted

In line with the evaluation survey, all the interview subjects expressed appreciation for the experience of learning in small groups. As noted earlier, in the course evaluation participants expressed only a narrow preference for heterogeneous groups. The overwhelming majority of interview subjects, however—nine out of ten—expressed a preference for heterogeneous groupings. For instance, one learner with audit experience who was in an heterogeneous group noted:

> I do like sharing, you know if I know something and I can help people, I really like doing that, so having the younger people in there kind of helped me do that. But it's not like I was way ahead of anybody, ask them questions too about stuff that they figured out.

And from the perspective from someone without audit engagement experience who was in a homogeneous group:

> I see the methodology of placing experienced level all together, I get it, but in the same breath I really think it would be great to have maybe someone's that's worked there for six months and then maybe a new hire and then maybe an intern or make new hire and intern equivalent to each other, that's fine….I find myself running across the table to talk to the other group cause it's got three people that have been on audit clients for a month and a half. So, unless they're going for something different, it would be a lot helpful for maybe the new hire/intern type people to have a more experienced individual there. Just so that you don't have to raise your hand a million times, you would have a quick reference.

The lone dissenter in the group of ten was a learner with no audit experience in a homogeneous group who appreciated that he was never left behind:

> So I thought that in that sense it was just us at the table and all were kind of like new at this, that it helped in the long run. Cause they had the same questions that I did, and I never felt like I was left behind when we were like going over stuff. Cause I think, especially the table that was to the right of me, that the stuff they were talking about and they would almost go complete their entire thing but we would still be half way through because we would have more questions and then

> so I thought in the long run that was probably the best fit if not I probably would have fell behind just a little bit.

A couple of subjects who were in heterogeneous groups did encounter feelings of falling behind other group members. One was an individual with audit experience who was chosen to be interviewed because he seemed less active in his group in terms of asking questions or giving answers. “Oh, that’s ‘cause they were ahead of me. I didn’t want to sit there and get in line and ask the instructor a whole bunch of questions they just answered….I didn’t want to waste their time.” He did not express any misgivings about being behind his group members, but to the extent that he was less active than the other members of his group it may have had an effect on his learning.

The other individual who talked about falling behind had no audit experience and was in a group with two learners with a relatively high amount of audit experience. She explained:

> One of the guys rushes through everything and gets everything done, doesn’t really consult with anybody until he’s done. And then he’s like, ‘oh, where you guys at? How you doing?’ And, I don’t like that….However, I do like to try to figure it out myself before I just ask. So that’s really why, I do sit there awhile and try to figure it out before I ask.

While she felt like she could not go as fast as her groupmates, she did not necessarily see this as a negative. She noted that one groupmate was more helpful than the other and that there was value in having peers nearby who can help her: “But it’s nice that I can go to them and talk to them if I have questions instead of asking…[an] instructor.” She expressed further that having knowledgeable peer resources would keep her from feeling like she and her group were uncomfortably monopolizing the instructors: “I like the mixed group probably better. That way you don’t feel like you’re hogging the instructor.”

Discomfort with falling behind had been a theme on course evaluations before the redesign. A few of the interviewees were chosen because they had taken the course before as interns. They were asked to contrast their experiences with their small groups. One subject, who was observed telling her classmates that the old course made her cry, described the old course in the interview:

> So I think that's why [the old course] was so hard because they just went over on the PowerPoint like OK, now you're going to get the bank statement or whatever and you're going to look at this, you're going to be looking for anything over your scope, this is what scope is and so anything over your scope you're going to ask for this and this and this and then you're going to do that. And then they would say OK now here's your example mock audit to do. And they like gave it to us and they had people who were going around to help us but it was like I think maybe, I'm sure because they didn't want to tell us too much, but it almost came across as, like, some of the questions that were just like learning questions that you would have to ask, it was like they wouldn't answer them. So it was almost like seeming mean on their part, you know? And then like you know my friends of course, I didn't want to be like, guys I have no idea what's going on. Cause you know, they all know it and it's like making sense to them and everything.

Another subject contrasted the old course and the new one this way:

> I think more the reason was because there was more opportunity for discussion in the earlier days of the [new] course. I think last year with the PowerPoint slides you're jammed full of so much information on the first couple of days that when you're sitting there, and you've had no experience on audits whatsoever, you sit there and you go to the case studies, and you're like, OK, have to think about what I remember from the two days when you didn't really have a chance to discuss it or talk over it, you've just had it grilled in your head from a PowerPoint lecture standpoint. So, you know, and you sit, obviously as anyone knows, you sit in a lecture for eight hours, you know once you get towards the end of the day not all that stuff's sticking at that point. So I think if you've had experience last year and the experienced groups were together, you've been on the experienced end of it, you were able to retain it, you know discuss with people in your groups, you knew where you were going, whereas if you were a new person, you couldn't remember all that information and you had no experience really to know where you were going. So I think that's why it made it easier was the small groups and the discussion.

The redesigned course specifically featured far more scaffolding than the old course. The hope of the design team was that the infusion of structure would keep advanced and novice learners somewhat more equitable in terms of pace. The interviews provided some evidence that the team was successful, that novices in the class were not falling far enough behind the advanced learners to reach the point of frustration.

Among the experienced learners interviewed, one theme was an appreciation for the learning aspect of tutoring less experienced learners.

> I actually liked being in the mixed group right now—I mean, there's pros and cons to each but I found myself thinking back to when I went on audits last year and having that information kind of trigger in my mind by the teaching aspect of it and using that you know for someone…who hadn't been on an audit before, explaining stuff to her, and I found the stuff that I went through last year was coming in my mind a lot more quickly because not only was I learning it in here but I was also taking the opportunity to teach it so you start to think about certain things, you thought about times on your audits last year, and all that relevant experience that you can bring to here.

Tutoring in his view has a valuable reflective aspect. According to another subject, being the experienced person in a group fielding questions can also be a reality check for you on how well you know the material.

> Having people like come to me to ask me questions just makes me realize, either realize what I know, and then when you have to like explain stuff, it's the whole thing of, when I would tutor accounting to students, I mean obviously if you're teaching something you have to know it because, so I think that helps. I think maybe if I hadn't been with people that I was explaining stuff to, I might not even have been able to develop the stuff that I knew as much cause I mean though I've done it before I haven't done it since last summer so it was like bringing it all back and like having to get really down to the real reasons and explaining it. So, I thought that was, I mean I got something out even if being with people who didn't have as much experience.

One subject pointed out that the experience of tutoring peers in the course may have real world impact:

> I'd rather be in a mixed group. And there's going to be downsides, but there's going to be upsides, too. Cause you're going to have like the people with more experience and they're gonna, it's going to help them more to show other people how to do it, cause management training's going to happen in three years, or in-charge training's going to happen in three years they might as well get used to showing people how to do something.

The ability to teach new staff the skills of auditing is important for auditors who "in charge" an engagement, and being a tutor in the course may help to begin building that ability. But, again, it should be acknowledged that despite the theme in the interviews of the value of peer teaching within groups, a narrow majority of experienced learners on the course evaluation preferred or would prefer homogeneous groupings.

Another theme that emerged was widespread acknowledgement that there was not a perfect solution in regards to experience grouping—that either approach entailed trade-offs. One subject noted:

> Because [the learners in a nearby experienced group] were—they were rolling like crazy, they're all going 100 miles an hour. And that's why I thought to myself, OK I see why they're segregating this [class], so the, yeah, there's definitely going to be weaknesses of both, and maybe splitting it up solves the one person running with it the entire time, if we could, I know that it's just pretty, and maybe that's the point. Maybe suffering through it is going to help you learn it. Cause I am, I've learned a lot, and having to suffer through that, granted it's taken us a lot longer, but, you know, maybe that's just the nature of the beast….Yeah, what you're trying to battle is crazy. If you had 500 people, and half of them were interns, and half of them like, and you could separate and do their own classes, that'd be fantastic, but you don't. It's going to be really hard to cater to so many different experience levels.

The same subject noted that range of experience is even more multi-faceted than it seems. Learners are not only separated by raw audit experience, but by knowledge differences stemming from differences in educational backgrounds:

> Just different from [academic undergraduate] program to program is huge. Going out to recruiting events and talking to people that are intellectual equals but in the same breath they're going through a program that emphasizes a different side of auditing. And there's concepts that they've never each heard of. And so you're battling that, as well.

He pointed out that one practical aspect of these differences is in the focus of the course. Learners who have been through recent, high quality audit courses are likely to have different conceptual needs than learners who are a couple of years removed from their last audit classes.

Chapter 5: Discussions and Recommendations

This chapter will interpret the findings presented thus far, note the limitations, and suggest implications for instructors and instructional designers as well as future areas for research. The following research question framed this study:

When professional, adult learners in problem-centered technical training courses are grouped heterogeneously and homogeneously by content experience, what is the effect on course effectiveness as measured by:

1. Learning gains
2. Problem-solving discourse
3. Learner satisfaction

This research was sparked by a real world problem. The sponsoring organization was running a high profile class that was not meeting its expectations, in part because the course lacked meaningful scaffolding built around its extensive use of ill-structured simulation. Compounding the problem was that small groups were not acting the way the organization hoped—more experienced learners did not appear to be successfully tutoring less experienced learners. The design team rebuilt the course from the ground up to include a more meaningful and scaffolded problem progression; however, the question remained about whether even with this more supportive course architecture learners should be grouped heterogeneously by experience, so that less experienced learners could benefit from the tutoring and insights of more experienced learners, or whether learners should be grouped homogeneously, keeping experienced auditors from dominating the problem allowing and allowing instructors to concentrate their efforts on the less experienced groups and assign enriching activities to more experienced groups.

Findings and Discussion

Relationship between group type and learning gains. This study was designed to compare learning gains between two types of small groups: homogeneous groups, those featuring only inexperienced or only experienced auditors, and heterogeneous groups, those featuring a mix of experienced and inexperienced auditors. Based on prior research indicating that homogeneous groups, at least when learners are grouped by

ability, tend to produce higher learning gains as learners advance in age toward adulthood, the prediction was that homogeneous groups would achieve higher learning gains.

In contrast, in this study heterogeneous groups outgained homogeneous groups, but only marginally, and the difference was not statistically significant. This lack of a significant difference suggests several possibilities. Among the possibilities is that the advantages and disadvantages of each type of grouping may more or less cancel each other out, leaving both types roughly equal at least in terms of learning gains.

It is also possible that adult learners at a professional working level adopt strategies that compensate for the disadvantages of each type of grouping. From classroom observations, for example, it was clear that some groups who were stuck would appeal to nearby groups for help despite instructions and reminders to work within one's group as much as possible. This behavior may suggest a high degree of self-regulation. While from observations of the classes and post-course interviews it appeared that most groups made an effort to treat their group as a team, there was certainly also instances observed where two small groups would temporarily form an alliance of sorts and essentially become one larger group. There were also some observed instances where learners who were in different groups but who were near each other would work together more closely than they would work with their own groups. As this was a real world setting, such freedom of movement was difficult to control. A laboratory setting could help prevent such crosspollinations, but application of those results to the real world would be problematic; it would not necessarily reflect how learners act in the real world.

Available evidence from prior offerings of the course suggest, however, that adult learners may not automatically have the coping skills to compensate for poor small group experiences. One reason this course was redesigned from the ground up was in response to numerous reports, from both observers and learners, that learners grew frustrated and shut down while operating in groups where they were not able to keep up with the most knowledgeable group members. However, it is also possible that the additional scaffolding added to the course had the effect of keeping learners close enough to mitigate this effect—that as long as the structure of the course provided sufficient

scaffolding to keep learners relatively close together, they were capable of using strategies to bridge the remaining gaps.

Another possibility is that the length of the course, forty hours, was a significant factor. Webb (1984) found that if groups were kept together over a long term, differences in achievement between homogeneous and heterogeneous groups eventually disappear. It is possible that with this target population, a week of concentrated learning may be enough to achieve this effect and that a shorter class may have shown different results. Note, though, that John Baer (2003) found in a study of college students, a population much closer to the present study, that differences between homogeneous and heterogeneous groups actually became stronger over the course of a semester in favor of homogeneous groups. To that extent, it is possible that a longer course in this study would also have shown different results.

By focusing on the nature of exchanges between learners, as opposed to motivations or social relationships, this study viewed small group learning through an cognitive elaboration perspective (O'Donnell & O'Kelly, 1994; Slavin, 1996). The guiding principles of this perspective are that learning in small groups is at least in part driven by the act of articulating mental models—that this articulation not only exposes gaps and forces refinement of the models, but the very act of anticipating a need to explain one's cognitive model initiates a process of internal rehearsal, elaboration, and exploration of anticipated conceptual gaps in the help-seeker, and this process itself is inherently valuable (Roschelle, 1992; Zajonc, 1960).

From the cognitive elaboration perspective, a lack of significant difference in learning gains between heterogeneous and homogeneous groups would indicate that learners in both types of groups were, on average, elaborating more or less equally in terms of quality and quantity of elaboration. In this study, elaboration was measured by monitoring the number of *explanations* sought and given, as Noreen Webb has shown across a number of studies (e.g., Webb, Welner, & Zuninga, 2001) that seeking out and attempting to supply meaningful descriptions of key concepts and principles correlates with achievement. Indeed, when compared across all the activities in the course,

explanation-seeking and explanation-giving percentages, as compared to other types of information exchange, were similar between the two types of groups.

Interestingly, when the comparison of explanation-focused strategies between the two types of groups was limited to ill-structured problem-solving activities, a gap between heterogeneous and homogeneous groups appeared to emerge, with heterogeneous groups showing a higher percentage of explanation-seeking, intended explanations, and successfully delivering correct explanations. However, variability was high, and restricting the data to ill-structured activities severely limited the statistical power of the comparison, rendering the differences statistically insignificant. More research is needed to discover if this trend was an anomaly or whether heterogeneous groupings for ill-structured problem solving really do encourage the seeking and granting of explanations.

Role of problem structure. While the impact of ill-structured problem solving on explanation-focused exchanges in heterogeneous groups was statistically insignificant, the analysis of the qualitative data did illuminate a strong significant difference in explanation-focus between groups solving relatively ill-structured problems versus groups solving relatively well-structured problems. This finding provides evidence in support of Merrill and Gilbert's assertion (2008) that problem-centered instruction inherently increases positive group interaction.

Merrill and Gilbert's distinction between problem-*based* instruction and problem-*centered* instruction may be an important factor. Before the redesign, the available evaluation evidence for *Starting Your Career* indicated that productive problem-solving dialog was often absent in groups. Merrill and Gilbert would likely see the old course as problem-based—learners were systematically taught various concepts and principles before being asked to put it all together in lengthy audit simulations. Merrill and Gilbert contend that such an arrangement discourages the formation of rich, applied mental models that can be shared and refined between peers. The redesigned course, however, was designed to be more problem-centered by building these mental models through a series of applied problems.

From a cognitive elaborative perspective, it is the analysis and refinement of these mental models through dialog that is the value of cooperative learning. Without a nascent mental model to analyze and refine, effective elaboration cannot take place. Lacking a knowledge of what questions to ask, help-seekers cannot meaningfully ask for explanations and peer help-seekers may not even know where to start in bridging the conceptual gap, encouraging them to provide shallow descriptions or just providing a procedural answer outright.

Merrill and Gilbert did not specifically address considerations of small group composition as it relates to ability or experience. It may just have been out of scope for their article, but perhaps the finding of no significant difference in this study in terms of learning gains reflects their emphasis on the central importance of a well-designed problem-solving task. Perhaps the effects of techniques such as ability grouping or reward modulation fade as the systematic presentation of ill-structured problems itself encourages productive cooperative dialog.

Relationship between experience spread and learning gains. If future studies support the strong positive—though in this study not significant—relationship between heterogeneous groupings and explanation-focused exchanges, then one would expect a strong advantage for heterogeneous groups in terms of learning gains. Such was not the case in this study, as members of heterogeneous groups had only a marginal, insignificant advantage.

One possible mitigating factor is that at some point, raw expertise may cease to be an advantage in helping novices. Jonassen (1997) noted that experts and novices have different schema for domain-specific problem solving—that the difference is not one of degree but is fundamental. He and Hernandez-Serrano (2002) asserted, “Expertise often results in a knowledge compilation that makes articulation of knowledge about the domain difficult for the expert. Experts are often unable to unpack and articulate the reasoning behind their problem-solving activities” (p. 71). If relative experts are unable to provide meaningful explanations to novices, then from the cognitive elaboration perspective the value of grouping heterogeneously would be severely diminished.

Fifteen audits separated the least experienced auditors in this study from the most experienced auditors. It is conceivable that somewhere along this gap can be found a point of diminishing or even negative returns. To test for this possibility, a second analysis of the quantitative data was performed, this time comparing experience spread in groups against learning gains. Such an analysis shows no significant correlation once pre-test scores are controlled for. The lack of a significant correlation may indicate that 15 engagements is not enough of an experience spread to create an insurmountable expertise gap for this target population.

Another possibility is that experience spread has a bifurcating effect, that some groups thrive when a highly experienced member is present to share experiences and others wither. This possibility is supported by the borderline significant finding ($p = .053$) that the relationship between explanation-seeking and experience spread is parabolic; as experience spread increases, the percentage of explanation-seeking questions tends to either increase or decrease as compared to groups with little spread.

This nearly significant effect in fact parallels a similar finding by Webb, Welner, and Zuniga (2001), who found consistent patterns with high ability learners in homogeneous groups but highly inconsistent results with high ability learners in heterogeneous groups. The researchers found that some high ability learners in heterogeneous groups supported explanation-focused exchanges by providing detailed explanations both in response to questions and in anticipation of questions. Other high ability learners did not exhibit these behaviors, however, and achieved significantly less on the post-test. The researchers, unfortunately, were not able to determine *why* these patterns developed.

Because the present study utilized Webb's frameworks for data collection and analysis of qualitative data, this study has the same explanatory limitations. The high-experience-spread, high-explanation-seeking groups in this study did contrast with the high-experience-spread, low-explanation-seeking groups, however, in at least one respect. Both of the high-spread, high-explanation-seeking groups were made up of two relatively experienced learners and one inexperienced learner, while all three of the high-spread, low-explanation-seeking groups were made up of one experienced individual and

two novices, preliminarily suggesting that putting novices with multiple experienced individuals increases the odds of forming a connection that encourages the sharing of explanations. One might expect this to be true if personality was a large factor in play, as it would double the odds that the novice would find a compatible personality in the set of experienced learners available. Webb (1982a, 1982b) started down this road of inquiry when she tracked level of introversion as a possible factor, but did not pursue this line of research.

Another possibility is that having two experienced individuals in a group might provide a positive foil of sorts for each other—that is, with two experienced individuals in the group, they would have each other to maintain a level-appropriate dialog with while at the same time supporting each other in providing explanations to the novice. There were certainly occasions where this appeared to be the case in this study, where the experienced learners collaborated on explanations to a novice, but the analysis in this study was not designed to explore this possibility systematically.

Another possibility is that a single experienced individual grouped with multiple less experienced individuals might feel more like a tutor than a student, which is perhaps demotivating, though this study uncovered no direct evidence of this possibility either way.

One other interesting facet of this comparison is that both novice learners in the high-spread, high-explanation-seeking groups indicated on the end-of-course evaluation that given a choice they would prefer a heterogeneous group. However, among the six relative novices in the high-spread, low-explanation-seeking groups, 83 percent—five out of six—indicated that they would have preferred a homogeneous group. That is striking considering that only nine of the remaining 27 learners (33 percent) who had less experience than the most experienced learner in their group indicated a preference for homogeneous groups.

These tentative findings would suggest that heterogeneous groups with multiple experienced individuals would be optimal. Clearly, in situations where experienced individuals are balanced roughly equally with novices, as was the case in this study, such a ratio is not feasible unless group size is increased to four individuals, two experienced

and two inexperienced—though increasing the group size might have negative impacts due to findings that smaller group sizes correlate with greater achievement (Lou et al., 1996; Wheelan, 2009).

Relationship between experience grouping and satisfaction. Course satisfaction scores were very high for this course both in comparison to the same course before the redesign and to other courses offered at the firm. In fact, one of the two classes that were part of this study was the highest scoring large course (20 or more respondents to a course evaluation) in the entire training year (out of 363 courses that met this minimum), and the other class in the study was also in the top ten percent.

Satisfaction scores were slightly higher for learners in heterogeneous groups than in homogeneous groups, but the difference was not significant. There may have been a ceiling effect in play since scores bumped up against the top of the scale.

When asked which type of group each learner would like to participate in, 57 percent of respondents said they would prefer to work in heterogeneous groups over homogeneous groups. Though learners do not necessarily make choices that are in their best interest instructionally (Clark & Mayer, 2002), that does not mean their choices should be automatically discounted, especially in the absence of compelling effectiveness data one direction or the other. A complicating factor here is that the results were not consistent across experience levels. Sixty-five percent of inexperienced learners expressed a preference for heterogeneous groups but only 47 percent of experienced learners made the same choice.

Limitations of the study. Several limitations of the study should be considered when interpreting the results.

***An imprecise definition of* experience**. One limitation of this study is that it defined *experience* crudely, as a function of the number of audit engagements. One charm of this definition is that it can be easily assessed; instructors can readily ascertain the approximate number of audit engagements a learner has been on at the beginning of a class during introductions. However, there are many nuances of experience that are not captured by the raw number of engagements. A learner in this study could have been on one or even multiple engagements where he or she filled a role that taught him or her

little about audit engagements—that auditor might have filled a merely administrative role, or been given little oversight and mentoring. Those learners would be considered experienced in this study but may be little better off than learners who had been on no audit engagements.

Further, practical experience may come from places other than audit engagements. One learner in this study with no formal audit engagement experience was a nontraditional student that had worked for years in an accounting department and thus had significant experience with audits from the other side. This learner was cast in the study as an inexperienced learner in a heterogeneous group, but in the transcripts and observations of the group he dominated the discussions and the explanation-giving, often offering his experiences dealing with auditors. While this was the only learner identified by the instructors and researcher as being obviously miscategorized under the label *inexperienced*, there may have been others. It is possible that a different or more comprehensive definition of *experience* would have led to different groupings and different results.

Measurement of learning gains. It is possible that the instrument used to measure learning gains did not accurately capture the level of learning that took place in the small groups. The pre-test was created by the researcher partnering with a subject matter expert with the goal of writing a test that reflected the stated learning objectives. The test was vetted by a third subject matter expert, piloted with two subjects similar to, though slightly more advanced than, the target population. The test was revised based on their feedback. A modified version was then created to become the post-test. It is possible that a test creation effort that involved significantly more piloting and application of test-creation expertise might have resulted in a materially different test that might have revealed significant differences in learning gains.

Another potential weakness of the instrument might be the form of the instrument itself. A multiple choice/short answer format may be insufficient for measuring learning from ill-structured problem solving. A more robust, though more difficult to administer, approach might have been to have participants independently complete simulations while being monitored by a trained observer. Likewise, monitoring participants on their first

jobs after the course might have provided insights into their abilities to solve real world, complex problems.

Lack of a robust problem-solving typology. The lack of a robust typology for pegging problem-solving activities along a continuum from well- to ill-structured makes the results from this and other problem-solving research difficult to generalize. Given the current state of problem-solving scholarship, it is difficult to express precisely how much less structured one task is than another, especially across content domains.

Possible weaknesses in instructional design related to problem solving. While the instruction featured in this study in many ways fits Merrill and Gilbert's specifications (2008) for optimal small group problem solving, it does not fit all of them. The course analyzed in this study was based on a cognitive apprenticeship model; Merrill and Gilbert's analysis was specific to instruction built around van Merriënboer's 4C/ID model of instructional design (van Merriënboer, 1997; van Merriënboer & Kirchner, 2007). The cognitive apprenticeship model as implemented in this study lacked the truly iterative progression of problem solving advocated by van Merriënboer and his colleagues. It is possible that a truly iterative progression of problems would have led to different results.

Another potential weakness in the instructional design of the course relates specifically to problem solving in small groups. Elizabeth Cohen (1994) in her analysis of problem solving in small groups drew a distinction between "group tasks," where group members effectively have to collaborate to solve problems, and routine learning tasks where learners collaborate on problems that could realistically be solved individually. Cohen was not necessarily advocating jigsaw-type activities, where information is artificially split up among group members, but rather she was arguing in favor of problem solving of sufficient complexity where exchanges of multiple points of view become essential for thoroughly mapping out solutions. To some extent, this complexity comes naturally as problems become less structured (Webb, Nemer, Chizhik, & Sugrue, 1998). However, there is no bright line in the literature beyond which a problem becomes sufficiently complex to move beyond routine learning and becomes a true group task. It is

possible that instruction featuring more complex problems than in this study might have returned different results.

Possible mismatch between observation and analysis instruments and the target population. The small group observation and analysis instruments utilized in this study were built on the extensive work of Noreen Webb and her colleagues. By choosing these instruments, the study made the assumption that meaningful achievement in learning groups would correlate to factors such as explanation-seeking and explanation-giving. However, Webb's target populations have all been much younger than the target population of this study. It is possible that her instruments would be insensitive to important factors that predict success in small group learning in adults. Further, her instruments have been used largely to analyze problem-solving discourse during well-constrained problem solving. The use of her instrument might mean that important markers in ill-structured problem-solving discourse were not coded during the analysis of this study's data. Finally, Webb tends to describe her instruments in ways that are specific to the problems being solved in her studies, forcing some adaptation in using her instruments in the present study. Adaptation involves interpretation, and it is possible the researcher may have misinterpreted her intent—for example, Webb may have drawn a different line between explanation-seeking and information-seeking.

One possibility for overcoming this limitation in future research might be the application of a more sophisticated framework for analyzing student discourse. One possibility is the framework suggested by van Boxtel and Roelofs (2001) which is built around the constructs of conceptual discourse, elaboration, co-construction, and the use of available tools.

Limited statistical power. The number of subjects in the study limited the power of some of the statistical analyses, especially when the results were broken out by experience range rather than binary category, and also when the relatively ill-structured problem-solving activities were isolated for analysis. A larger dataset might have confirmed or disconfirmed some of the near-significant findings of this study.

Lack of a perfectly controllable laboratory environment. This research was conducted in a live classroom instead of more of a easily-controlled laboratory setting. As

a consequence, some deviation from the research protocols were observed. While groups were periodically reminded to stick together in their teams, which they clearly did for the most part, there were also observed instances of groups appealing to each other for help or of an individual from one group breaking off temporarily to help an individual in another group. While these moments were the exception rather than the rule, it is possible that crosspollination between groups was sufficient to mask results that would have been uncovered in a more tightly-controlled—but therefore perhaps less realistic—setting.

Middle learner effect. One limitation of the design of this study is that it failed to explore systematically any potential effects of the middle person in the group in terms of experience. Not every group had someone in the middle in terms of experience, but some did. Webb (1991) identified the middle members of small groups, neither the highest nor the lowest performers, as being problematic because they do not give as many explanations as higher ability members nor do they receive as many as lower ability members. Systematically controlling for the effect of middle learners may have led to useful insights in interpreting the rest of the data.

Limited control of instructor interaction with groups. Part of the argument in favor of homogeneous groupings is that it theoretically frees instructors to assign enriching activities to faster groups and spend more time with struggling groups. To that end, this course design included "Going Beyond" activities that could be assigned by instructors to faster groups. However, there was no evidence from the course observations that they were used consistently across classrooms.

Likewise, there was little systematic control over how much time instructors spent with different group types during small group activities. Based on earlier trials of grouping learners heterogeneously or homogeneously, instructors at this firm had expressed appreciation for homogeneous groupings because they could spend more time with the groups that needed it most. However, the analysis for this study revealed no significant relationships either between group type and prevalence of teacher interventions nor between teacher interventions and learning gains. It is possible that greater control over instructor direction in these regards may have produced different results.

Limited generalizability. While this research was intended to expand the cooperative learning research to post-college adults, a population not generally addressed in this body of research, in fact the population in this study may represent more of a transition population than a true professional adult population. The subjects in this study were generally only a few months removed from completing their accounting degrees, with a majority of subjects embarking on their first career. Thus, most subjects in the study might respond more like college students than professional, adult learners who are further removed from college. In other words, it is possible that participating in a profession gives one a different perspective, level of maturity, or set of internal motivations that could alter his or her reactions to cooperative learning. Professionals advanced enough in their careers to be routinely mentoring and training others, for instance, may more readily embrace heterogeneous groups, or perhaps the opposite if they feel resentment over doing the trainer's job.

Likewise, this study may lack generalizability downward to college-age populations. Being in a professional training course, where the relationship between learning and paycheck is clear and present, may create a much different cooperative learning dynamic than in college where the relationship is much more abstract.

Implications and Recommendations for Future Research

Recommendations for future research are presented below based on the findings and limitations of this study.

Development of a robust problem-solving typology. The generalizability of this and other problem-solving research will be hampered by the lack of a robust problem-solving framework that can help researchers and instructional designers compare results across domains and problem-solving activities. Jonassen (1997) started this process by highlighting attributes that separate well-structured from ill-structured problem solving, but considerable work remains to define precisely the continuum in a way that can be reliably applied to categorize a given problem-solving task.

Development of a unified model of expertise. This study relied on the concept of experience as a predictor of success in problem solving. *Experience*, however, is only one facet of *expertise*—an important facet, to be sure (Shin, Jonassen, & McGee, 2003),

but not the complete story. Domain knowledge such as might come from research, training, or academic learning, is certainly another major factor (Ertmer et al., 2008), as is intelligence, creativity, and other innate abilities (Sternberg, 1998).

Experts can reasonably be defined as those with the greatest ability to solve problems in their domains. Currently, however, there is no robust model for describing expertise systematically—in other words, there is no system that would allow a comparison of individuals, given their level of domain experience and knowledge and other factors, to predict who is most likely to be able to solve arbitrary problems in that domain.

In the present study, "number of audit engagements" was used as a shorthand for expertise, but that shorthand devalues analogous experiences such as those held by the learner with a real world accounting background who has experienced auditing from the viewpoint of an accountant. It also devalues academic experience by assuming that someone with excellent technical understanding from recent college experiences is less of an expert than someone who may have less academic expertise in the subject but has been out on real world audit engagements.

As noted earlier, this shorthand is appealing because it is practical for use in technical training classrooms. A system that involves development and implementation of pre-tests is probably a non-starter for many real world applications, even if the potential payoff is large. But is the shorthand close enough to still make it useful in technical training situations? If not, can it be tweaked to include elements like analogous experiences? Or is any shorthand going to be so crude that it cannot succeed? These are all open research questions.

Clarification of the relationship between experience spread and explanation-seeking. One of the near significant findings of this research is that explanation-seeking may have a parabolic relationship with experience spread in a small group. That possibility is interesting because it may indicate that as expertise increases, some factors come into play that encourage some help-seekers to ask for explanations while discouraging other help-seekers from doing the same. Could it be that some learners had reached a level of expertise that worked against them in providing explanations to

novices? Or perhaps there were personality factors that became magnified as the spread of expertise grows. To tease out this effect would require further research, first validating the effect and then exploring the nature of the relationship.

More generally, additional research that qualitatively explores the nature of explanation-seeking might help to explain why some learners sought or volunteered explanations and why others did not. What techniques are used by explanation-givers to make explanations understandable by help-seekers? Is the important difference really the division between explanation and information or are there finer gradations that are equally important? What causes help-seekers to persist in the face of unhelpful explanations? Or, what causes them to retreat from explanation-seeking? Given the uncertainty, a grounded theory approach, with its focus on generating theory (Patton, 2002), might be the most appropriate vehicle for future research.

Extending other aspects of Webb's research. Webb and her colleagues have shown that applying explanations independently to novel, relatively well-structured problems correlates with increased achievement (Webb, Troper, & Fall, 1995). This aspect of small group problem solving was out of scope for this research effort. Tracking whether learners generalized explanations independently to novel problems would have necessitated a research design where individual learners were tracked across problems. Thus, the question of whether independent application of explanations to novel ill-structured problems remains an open area of research.

Another thread in Webb's research is the role of the teacher in modeling explanation-giving behavior (Webb et al., 2008, 2009). This thread was also unexplored in this research, as instructors received no special instructions or training in modeling the sharing of explanations. Further research would be required to discover whether, with modeling, the most experienced learners in the class can be turned into more effective tutors.

Implications for Design and Instruction

In terms of learning gains and satisfaction, this study found no significant differences between heterogeneous and homogeneous experience groups. However, even a lack of significant difference carries instructional implications.

Implications for instructors. The key driver of this research was to provide guidance to instructors of technical, relatively ill-structured content to professional, adult learners in how to group learners when the relative range of experience is large. The results of this study indicate that, at least in terms of learning gains, discourse, and satisfaction, it may not matter.

Absent a preference one way or the other on the part of the instructor, one way to resolve the issue would be to put weight on the finding that 57 percent of learners in this study that if they had the choice they would choose heterogeneous groupings. However, less than half of learners with experience made that same choice; therefore, choosing heterogeneous groupings may put a majority of experienced learners into groups they would not choose on their own. Instructors could leave the choice up to learners, each selecting the type of group he or she would prefer. However, this option presents a number of difficulties: not enough experienced learners may choose heterogeneous groups, the logistics may take too much time to sort out, prior to a class learners may simply not know which one to choose, and the possibility of social pressure driving learners with experience to choose heterogeneous groups when they might prefer homogeneous groups.

Depending on the instructor or the course, pacing may be a factor to consider. Heterogeneous groups are, logically, more likely to work at approximately the same speed than homogeneous groups, reducing classroom management difficulties. On the other hand, by virtue of working at a range of speeds, homogeneous groupings provide instructors with an opportunity to challenge faster groups and provide more attention to struggling groups. Instructors who choose to go this direction should be ready for this split so advanced groups do not end up spending considerable time being idle and so struggling groups get appropriate instructor attention. To this extent, one compromise position might be for experienced instructors to group learners homogeneously to take advantage of the possibilities but for inexperienced instructors—many times in technical training courses instructors are subject matter experts with little training in instruction—to group learners heterogeneously to avoid the additional group management difficulties.

Another compromise position might be suggested by one of the difficulties in the current study. Instructors often do not have control over the learning space. In this study, tables were set up for six learners per table; given space in the room and running power and network cables, rearrangement would not have been trivial. Therefore, groups were arranged two per table. While groups for the most part respected the integrity of their own small group, there was definitely some crosspollination going on as groups would sometimes appeal to the group across the table rather than to the instructors for help. One of the interviewees in this study noted their use of this strategy—first try to work it out together, then appeal to the other group at the table. Instructors may be able to leverage this idea by pairing homogeneous inexperienced groups with homogeneous experienced groups at the same table. One danger here would be that groups of three might for all practical purposes merge into groups of six learners, well beyond the optimal group size of three to four learners (Lou et al., 1996).

The good news is that it appears, at least with adult professionals engaged in relatively ill-structured problem-based technical training, that instructors cannot go too far wrong no matter which way they decide to group learners. The lack of significant differences between group types in terms of learning and satisfaction suggests that this target population is able to adapt, on average, to either group type. That said, the success of the redesigned course, where groups were formed intentionally, versus the original course, where groups were formed without consideration of experience levels, may indicate that instructors would be well-advised to make an intentional choice of a grouping strategy.

Implications for instructional designers. The dramatic difference between explanation-focused exchanges between small groups solving relatively well-structured problems and those solving relatively ill-structured problems in this study is evidence in support of Merrill and Gilbert's assertion (2008) that the nature of the problems that learners are asked to solve is a key to positive small group functioning. In this study, groups observed solving relatively ill-structured problems engaged in a significantly higher percentage of explanation-seeking and explanation-sharing.

However, the evidence supporting Merrill and Gilbert is only indirect as this study did not directly compare learning gains between groups solving ill-structured problems and groups solving well-structured problems. In addition, this study revealed no significant differences in learning gains between groups engaging in a relatively high percentage of explanation-focused exchanges and those engaging in a relatively low percentage. That said, the large effect size in favor of explanation-seeking exchanges in relatively ill-structured problem solving would suggest that instructional designers interesting in supporting small group learning should focus primarily on the underlying problem-solving structure of the instruction as the key to positive group functioning.

Absent a strong recommendation for grouping learners either homogeneously or heterogeneously, where feasible instructional designers should build in support for both approaches. Since homogeneous groups may move at different speeds, designers could create, for instance, accelerated activities for faster groups to tackle when they have finished problem solving (Kulik, 1992). Likewise, designers should consider building in support for the needs of groups of novices. At a minimum this support should include advisements to instructors to work closely with these groups, as they are more likely to struggle. When instructors are working with these groups, it is important that they move learners toward conceptual understanding rather than toward the answers (Webb et al., 2008; Gilles & Khan, 2009). Given the extensive amount of training generally given to instructors in studies focused on guiding small groups toward conceptual understanding (e.g., Nattiv, 1994)—and these studies are with credentialed teachers as opposed to volunteer technical experts—merely giving tips in a leader guide and a train-the-trainer session may not be sufficient. Therefore, designers should work toward embedding scaffolds in the problem solving that encourage learners to articulate and refine their conceptual models.

For instance, in *Starting Your Career,* one kind of activity utilized a number of times was giving learners a narrative of how an auditor solved a particular problem. Learners were asked to annotate the narrative to describe why a particular step taken by the auditor was important, or why it was not. Learners were asked to do similar activities while observing live and prerecorded models, and then compare notes in their small

groups (Merrill, 2002; Collins, Brown, and Holum, 1991). In line with the cognitive elaboration model, these kinds of activities encourage articulation of conceptual understandings no matter the group type. Designers need to try and find the line of providing enough structure to encourage explanation swapping without imposing so much structure that they smother the very dialogue they are trying to encourage (Cohen, 1994).

The research in this study provides some evidence in support of Webb's findings (Webb et al., 1998; Webb, Welner, & Zuniga, 2001) that explanation focus in heterogeneous groups is highly variable compared to homogeneous groups, perhaps even bifurcated. Therefore, to support instructors who utilize heterogeneous grouping, instructional designers should seek out activities that encourage learners with relatively high expertise to share explanations at an appropriate level and to encourage learners with relatively low expertise to be persistent in seeking out explanations. The trick, naturally, is resolving exactly how to achieve this effect consistently. It is possible that techniques designed to provide structured roles may be effective, such as reciprocal teaching (Palincsar & Brown, 1984) or training learners in seeking explanations (King, 1992). However, no research exists validating these techniques with adult, professional learners, much less in ill-structured technical training situations.

Instructional designers may be well advised, then, to follow Merrill and Gilbert's advice (2008) to apply Merrill's first principles (Merrill, 2002) to small group learning to encourage explanation-focused dialog. Following Merrill's activation principle, for example, encourages designers to build in activities where learners are encouraged to describe their prior experiences and how they solved problems. Application of the integration principle would similarly encourage learners to explain how they will apply strategies learned in the course to the real world.

Resetting the question. One potential advantage of heterogeneous grouping is that experienced learners can provide wisdom or support to novices in ways that have value but are difficult to measure. By relating their experiences—"cases" in the language of case-based reasoning (Kolodner, 1997)—learners may be able to build mental models that do not necessarily lead to higher post-test scores but that do help them apply their

knowledge more readily in the real world. Hearing real world stories may be reassuring. Hearing the stories of those who have "been there" recently enough to remember what it feels like to go on that first job may also serve an almost ceremonial, initiation function (Jonassen & Hernandez-Serrano, 2002). Further, the act of providing mentoring in small groups might be valuable in the long term for experienced learners who will quickly find such skills valuable on the job as new employees are hired and they find themselves in the roles of mentor and on-the-job trainer.

To that extent, perhaps asking which type of group is superior for learning is the wrong question. Perhaps the question of interest should be how to harness the natural advantages of heterogeneous groupings. To answer this question would involve finding ways to ensure that learners with greater expertise feel that their experiences helping novices was worthwhile. Answering this question would also involve teasing out the variability in heterogeneous groupings found both in this study and in Webb's research (Webb, Welner, & Zuniga, 2001) with an eye toward putting in place supports that encourage learners to strive toward the positive end of that variability.

This research was designed to solve a real world problem, but the study is also a step towards creating a body of literature that helps instructional designers assist subject matter experts to create technical training courses that leverage the existing expertise of learners in those courses. Much work lies ahead in creating a deep research base to guide subject matter experts who teach technical topics in harnessing the advantages of cooperative learning.

Personal reflections. By all accounts, the redesign of *Starting Your Career* at this firm was a great success story. Instead of languishing in the bottom third of the course catalog in terms of satisfaction, the course was one of the top performers offered by the firm. The tone of the comments on the evaluations has changed and are far more positive, especially in terms of cooperative group work. Given the results discussed throughout this paper, the greater satisfaction with cooperative learning appears not to be driven by type of learning group (homogeneous or heterogeneous), but probably by the greater level of scaffolding offered by the course, chunking out the information so that experienced learners cannot simply race ahead of their groupmates and providing a

structure for them to explain their conceptual understandings, and by the use of small groups of three instead of relatively large groups of six.

The firm has updated the course materials to not recommend a particular grouping strategy but rather to encourage instructors to choose one strategy or the other based on their goals and comfort levels.

In the meantime, I intend to continue to explore ways to unlock the potential power of heterogeneous groupings. Experienced learners have real world stories that can prepare, and hopefully reassure, newcomers about their career journey and cultural norms in the firm. The nearly significant results in this study suggesting a polarizing effect as diversity in the group grows—high diversity appears to either increase the depth of problem solving dialog or decrease it, a result that parallels those found by Webb (Webb, Welner, & Zuniga, 2001). No research to date has sought to untangle this effect. Are there personality factors in play? Are there types of experience or expertise that transfer well in small groups? Are there scaffolds or other kinds of instructional techniques that can eliminate this variability? I look forward to exploring these questions.

References

Ackerman, E. (2001). Piaget's constructivism, Papert's constructionism: What's the difference? *Future of Learning Group Publication*. Retrieved from http://www.learning.media.mit.edu/content/publications/EA.Piaget%20_%20Papert.pdf

AICPA. (2005). SAS No. 103. Retrieved June 4, 2009, from http://ebpaqc.aicpa.org/Resources/Accounting+and+Auditing+Resource+Centers/SAS+No.+103.htm

Alejandro, J. (2001). Understanding professionals' reasons for participating in continuing professional education. *Adult Learning*, *12*(1), 15.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., et al. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition* (2nd ed.). Boston, MA: Allyn & Bacon.

Baer, J. (2003). Grouping and achievement in cooperative learning. *College Teaching*, *51*(4), 169-174.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*(3), 75-78.

Bandura, A. (2004). Swimming against the mainstream: The early years from chilly tributary to transformative mainstream. *Behaviour Research and Therapy*, *42*(6), 613-630. doi:10.1016/j.brat.2004.02.001

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*(2), 164-180. doi:10.1111/j.1745-6916.2006.00011.x

Bassi, L. J., Benson, G., & Cheney, S. (1996). The top ten trends. *Training & Development*, *50*(11), 27.

Bockarie, A. (2002). The potential of Vygotsky's contributions to our understanding of cognitive apprenticeship as a process of development in adult vocational and technical education. *Journal of Career and Technical Education*, *19*(1), 47-66.

Bowling, N. A. (2007). Is the job satisfaction-job performance relationship spurious? A meta-analytic examination. *Journal of Vocational Behavior*, *71*(2), 167-185.

van Boxtel, C., & Roelofs, E. (2001). Investigating the quality of student discourse: what constitutes a productive student discourse? *Journal of Classroom Interaction*, *36*(2/1), 55-62.

Brewer, S., & Klein, J. D. (2006). Type of positive interdependence and affiliation motive in an asynchronous, collaborative learning environment. *Educational Technology Research and Development*, *54*(4), 331-354.

Brown, V. R., & Paulus, P. B. (2002). Making group brainstorming more effective: Recommendations from an associative memory perspective. *Current Directions in Psychological Science*, *11*(6), 208-212.

Carter, G., & Jones, M. G. (1994). Relationship between ability-paired interactions and the development of fifth graders' concepts of balance. *Journal of Research in Science Teaching*, *31*(8), 847-56.

Cheng, R. W., Lam, S., & Chan, J. C. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning. *British Journal of Educational Psychology*, *78*, 205-221. doi:10.1348/000709907X218160

Chiang, C. F., Back, K. J., & Canter, D. D. (2005). The impact of employee training on job satisfaction and intention to stay in the hotel industry. *Journal of Human Resources in Hospitality & Tourism*, *4*(2), 99–118.

Chinn, C. A., O'Donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *Journal of Experimental Education*, *69*(1), 77.

Clark, R. C., & Mayer, R. E. (2002). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (1st ed.). San Francisco, CA: Pfeiffer.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*(1), 1-35. doi:10.3102/00346543064001001

Cohen, E. G., & Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal*, *32*(1), 99-120. doi:10.2307/1163215

Cohen, E. G., Lotan, R. A., Abram, P. L., Scarloss, B. A., & Schultz, S. E. (2002). Can groups learn? *Teachers College Record*, *104*(6), 1045.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making things visible. *American Educator: The Professional Journal of the American Federation of Teachers*, *15*(3), 6-11,38-46.

Crook, C. (1998). Children as computer users: The case of collaborative learning. *Computers & Education*, *30*(3-4), 237-247. doi:10.1016/S0360-1315(97)00067-5

Dennen, V. P., & Burner, K. J. (2007). The cognitive apprenticeship model in educational practice. In *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 425-439). New York, NY: Routledge.

Didierjean, A. (2003). Is case-based reasoning a source of knowledge generalisation? *European Journal of Cognitive Psychology*, *15*(3), 435.

Dorval, B., Eckerman, C. O., & Ervin-Tripp, S. (1984). Developmental trends in the quality of conversation achieved by small groups of acquainted peers. *Monographs of the Society for Research in Child Development*, *49*(2), 1-91. doi:10.2307/1165872

Ertmer, P. A., Stepich, D. A., York, C. S., Stickman, A., Wu, X., Zurek, S., & Goktas, Y. (2008). How instructional design experts use knowledge and experience to solve ill-structured problems. *Performance Improvement Quarterly*, *21*(1), 17-42.

Ferguson, W., Bareiss, R., Birnbaum, L., & Osgood, R. (1992). ASK systems: an approach to the realization of story-based teachers. *Journal of the Learning Sciences*, *2*(1), 95.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: Sage Publications Ltd.

Fişek, M. H., Berger, J., & Norman, R. Z. (1991). Participation in heterogeneous and homogeneous groups: A theoretical integration. *The American Journal of Sociology*, *97*(1), 114-142. doi:10.2307/2781640

Flavell, J. H. (1996). Piaget's legacy. *Psychological Science*, *7*(4), 200-203.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal*, *35*(2), 227-267. doi:10.2307/1163424

Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research & Development*, *52*(2), 5-22.

Genberg, V. (1992). Patterns and organizing perspectives: A view of expertise. *Teaching and Teacher Education*, *8*(5-6), 485-495. doi:10.1016/0742-051X(92)90053-6

Gillies, R. M., & Khan, A. (2009). Promoting reasoned argumentation, problem-solving and learning during small-group work. *Cambridge Journal of Education*, *39*(1), 7. doi:10.1080/03057640802701945

Hooper, S. (1992). Cooperative learning and computer-based instruction. *Educational Technology Research and Development*, *40*(3), 21-38. doi:10.1007/BF02296840

Hooper, S., Temiyakarn, C., & Williams, M. (1993). The effects of cooperative learning and learner control on high- and average-ability students. *Educational Technology Research and Development*, *41*(2), 5-18. doi:10.1007/BF02297309

Jensen, K. (2007). The desire to learn: An analysis of knowledge-seeking practices among professionals. *Oxford Review of Education*, *33*(4), 489. doi:10.1080/03054980701476055

Johnson, D. W., & Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs*, *131*(4), 285-358.

Johnson, D. W., & Johnson, R. T. (2007). Cooperation and the use of technology. In *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 401-423). New York, NY: Routledge.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1988). *Cooperation in the classroom* (Revised.). Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R. T., Johnson, J., & Anderson, D. (1976). Effects of cooperative versus individualized instruction on student prosocial behavior, attitudes toward learning, and achievement. *Journal of Educational Psychology*, *68*(4), 446-452. doi:10.1037/0022-0663.68.4.446

Johnson, D. W., Johnson, R. T., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. *Review of Educational Research*, *53*(1), 5-54.

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*(1), 15-29.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Methods of cooperative learning: What can we prove works. Retrieved April 4, 2009, from http://www.co-operation.org/pages/cl-methods.html

Jonassen, D., Strobel, J., & Lee, C. B. (2006). Everyday problem solving in engineering: Lessons for engineering educators. *Journal of Engineering Education*, *95*(2), 139.

Jonassen, D. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65-94. doi:10.1007/BF02299613

Jonassen, D. H., & Hernandez-Serrano, J. (2002). Case-based reasoning and instructional design: Using stories to support problem solving. *Educational Technology Research and Development*, *50*(2), 65-77.

Jones, M. G., & Carter, G. (1994). Verbal and nonverbal behavior of ability-grouped dyads. *Journal of Research in Science Teaching*, *31*(6), 603-619. doi:10.1002/tea.3660310604

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: a qualitative and quantitative review. *Psychological Bulletin*, *127*(3), 376-407.

Kalkhoff, W., & Thye, S. R. (2006). Expectation states theory and research: New observations from meta-analysis. *Sociological Methods Research*, *35*(2), 219-249. doi:10.1177/0049124106290311

Kerka, S. (2002). Teaching adults: Is it different? Myths and realities. For full text: http://www.ericacve.org/pubs.asp. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED468614

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, *27*(1), 111-126.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*(1), 134-152. doi:10.1037/0022-0663.90.1.134

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Kohn, A. (2004). *What does it mean to be well educated? And other essays on standards, grading, and other follies*. Boston, MA: Beacon Press.

Koppenhaver, G. D., & Shrader, C. B. (2003). Structuring the classroom for performance: cooperative learning with instructor-assigned teams. *Decision Sciences Journal of Innovative Education*, *1*(1), 1-21. doi:10.1111/1540-5915.00002

Kuhn, D. (1972). Mechanisms of change in the development of cognitive structures. *Child Development*, *43*(3), 833-844. doi:10.1111/1467-8624.ep12114386

Kulik, J. A. (1992). An analysis of the research on ability grouping: Historical and contemporary perspectives. National Research Center on the Gifted and Talented, The University of Connecticut, 362 Fairfield Rd, U-7, Storrs, CT 06269-2007. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED350777

Kumplainen, K., & Kaartinen, S. (2003). The interpersonal dynamics of collaborative reasoning in peer interactive dyads. *Journal of Experimental Education, 71*(4), 333-370.

Lancaster, K. A. S., & Strand, C. A. (2001). Using the team-learning model in a managerial accounting class: An experiment in cooperative learning. *Issues in Accounting Education*, *16*(4), 549.

Lawrenz, F., & Munch, T. W. (1985). Aptitude treatment effects of laboratory grouping method for students of differing reasoning ability. *Journal of Research in Science Teaching*, *22*(3), 279-287. doi:10.1002/tea.3660220309

Lent, R. W., Schmidt, J., & Schmidt, L. (2006). Collective efficacy beliefs in student work teams: Relation to self-efficacy, cohesion, and performance. *Journal of Vocational Behavior*, *68*(1), 73-84. doi:10.1016/j.jvb.2005.04.001

Levitt, S. D., & Dubner, S. J. (2009). *Superfreakonomics: Global cooling, patriotic prostitutes, and why suicide bombers should buy life insurance* (First Edition.). New York, NY: William Morrow.

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *Journal of Educational Research, 94*(2), 101.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423-458. doi:10.2307/1170650

Maddux, J. E. (1999). The collective construction of collective efficacy: Comment on Paskevich, Brawley, Dorsch, and Widmeyer (1999). *Group Dynamics, 3*, 223-226.

Martin, J. (2004). Self-regulated learning, social cognitive theory, and agency. *Educational Psychologist*, *39*(2), 135-145.

Matusov, E., & Hayes, R. (2000). Sociocultural critique of Piaget and Vygotsky. *New Ideas in Psychology*, *18*(2-3), 215-239.

McAuliffe, T. J., & Dembo, M. H. (1994). Status rules of behavior in scenarios of peer learning. *Journal of Educational Psychology*, *86*(2), 163-72.

Meloth, M. S., & Deering, P. D. (1994). Task talk and task awareness under different cooperative learning conditions. *American Educational Research Journal*, *31*(1), 138-165. doi:10.2307/1163270

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, *25*(1), 95-111. doi:10.2307/1501934

Merriënboer, J. J. G. V. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Pubns.

Merriënboer, J. J. V., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Lawrence Erlbaum.

Merrill, M. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43-59. doi:10.1007/BF02505024

Merrill, M. D., & Gilbert, C. G. (2008). Effective peer interaction in a problem-centered instructional strategy. *Distance Education*, *29*(2), 199-207.

Messier, W., Glover, S., & Prawitt, D. (2006). *Auditing and assurance services* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Mulryan, C. M. (1995). Fifth and sixth graders' involvement and participation in cooperative small groups in mathematics. *Elementary School Journal*, *95*(4), 297-310.

Murtonen, M., Olkinuora, E., Palonen, T., Hakkarainen, K., & Lehtinen, E. (2008). Motivational orientations in work. *International Journal of Educational Research*, *47*(4), 213-222. doi:10.1016/j.ijer.2008.07.007

Nattiv, A. (1994). Helping behaviors and math achievement gain of students using cooperative learning. *Elementary School Journal*, *94*(3), 285.

Newbern, D., Dansereau, D. F., Patterson, M. F., & Wallace, D. S. (1994). Toward a science of cooperation. Presented at the American Educational Research Association Annual Convention, New Orleans, LA. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED367713

O'Donnell, A. M., & O'Kelly, J. (1994). Learning from peers: Beyond the rhethoric of positive results. *Educational Psychology Review*, *6*(4), 321.

Oetzel, J. G. (2001). Self-construals, communication processes, and group outcomes in homogeneous and heterogeneous groups. *Small Group Research*, *32*(1), 19-54. doi:10.1177/104649640103200102

Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist*, *35*(4), 227-241.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117-175. doi:10.2307/3233567

Paskevich, D. M., Brawley, L. R., Dorsch, K. D., & Widmeyer, W. N. (1999). Relationship between collective efficacy and team cohesion: Conceptual and measurement issues. *Group Dynamics*, *3*, 210-222.

Patton, M. Q. (2001). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Peterson, P. L., Janicki, T. C., & Swing, S. R. (1981). Ability × treatment interaction effects on children's learning in large-group and small-group approaches. *American Educational Research Journal*, *18*(4), 453-473. doi:10.2307/1162633

Pressley, M., Symons, S., McDaniel, M., Snyder, B., & Turnure, J. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology*, *80*(3), 268-278.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*(1), 91-109.

Ravenscroft, S. P., Buckless, F. A., & Hassall, T. (1999). Cooperative learning-A literature guide. *Accounting Education*, *8*(2), 163-176.

Reynolds, C. R., & Fletcher-Janzen, E. (2009). *Handbook of clinical child neuropsychology* (3rd ed.). New York, NY: Springer.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155-169.

Robles, H. J. (1998). *Andragogy, the adult learner and faculty as learners*. Retrieved from

http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED426740

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

Ross, J. A., & Cousins, J. B. (1995). Impact of explanation seeking on student achievement and attitudes. *Journal of Educational Research*, *89*(2), 109.

Saleh, M., Lazonder, A. W., & Jong, T. D. (2007). Structuring collaboration in mixed-ability groups to promote verbal interaction, learning, and motivation of average-ability students. *Contemporary Educational Psychology*, *32*(3), 314-331. doi:10.1016/j.cedpsych.2006.05.001

Schullery, N. M., & Schullery, S. E. (2006). Are heterogeneous or homogeneous groups more beneficial to students? *Journal of Management Education*, *30*(4), 542-556.

Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*, *40*(1), 6-33. doi:10.1002/tea.10058

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43-69. doi:10.1006/ceps.1996.0004

Smith, H. W. (1971). Interaction process in small groups of varied ages. Presented at the American Sociological Association Convention, Denver, CO. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED056338

Song, H., & Grabowski, B. (2006). Stimulating intrinsic motivation for problem solving using goal-oriented contexts and peer group composition. *Educational Technology Research & Development*, *54*(5), 445-466. doi:10.1007/s11423-006-0128-6

St. Clair, R. (2002). Andragogy revisited: Theory for the 21st century? Myths and realities. For full text: http://www.ericacve.org/pubs.asp. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED468612

Stapleton, J. L. (2007). Effects of team composition on problem solving: An empirical investigation of the assembly effect. *Delta Pi Epsilon Journal*, *49*(2), 94-108.

Sternberg, R. J. S. (1998). *Handbook of creativity* (1st ed.). New York, NY: Cambridge University Press.

Taylor, K., Marienau, C., & Fiddler, M. (2000). *Developing adult learners: Strategies for teachers and trainers* (1st ed.). San Francisco, CA: Jossey-Bass.

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

Watson, S. B., & Marshall, J. E. (1995). Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course. *Journal of Research in Science Teaching*, *32*(3), 291-99.

Webb, N. M., Welner, M., & Zuniga, S. (2001). *Short circuits or superconductors? Examining factors that encourage or undermine group learning and collaboration among high-ability students*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles; US Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center.

Webb, N. M. (1982a). Group composition, group interaction, and achievement in cooperative small groups. *Journal of Educational Psychology*, *74*(4), 475-84.

Webb, N. M. (1982b). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology*, *74*(5), 642-55.

Webb, N. M. (1984). Stability of small group interaction and achievement over time. *Journal of Educational Psychology*, *76*(2), 211-24.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, *22*(5), 366-89.

Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). 'Explain to your partner': teachers' instructional practices and students' dialogue in small groups. *Cambridge Journal of Education*, *39*(1), 49. doi:10.1080/03057640802701986

Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D., & Battey, D. (2008). The role of teacher instructional practices in student collaboration. *Contemporary Educational Psychology*, *33*(3), 360-381. doi:10.1016/j.cedpsych.2008.05.003

Webb, N. M., & Mastergeorge, A. (2003a). Promoting Effective Helping Behavior in Peer-Directed Groups. *International Journal of Educational Research*, *39*(1), 73-97.

Webb, N. M., & Mastergeorge, A. M. (2003b). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction, 21*(4), 361-428.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607-651. doi:10.2307/1163461

Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *Journal of the Learning Sciences, 15*(1), 63-119.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406-23.

Wheelan, S. A. (2009). Group size, group development, and group productivity. *Small Group Research, 40*(2), 247-262. doi:10.1177/1046496408328703

Winfrey, E. C. (1999). Kirkpatrick's four levels of evaluation. In *Encyclopedia of Educational Technology*. Retrieved from http://edweb.sdsu.edu/eet/articles/k4levels/index.htm

Wood, E., Pressley, M., & Winne, P. (1990). Elaborative interrogation effects on children's learning of factual content. *Journal of Educational Psychology, 82*(4), 741-48.

Zajonc, R. B. (1960). The process of cognitive tuning in communication. *Journal of Abnormal and Social Psychology, 61*, 159-167.

Appendix A: Semi-Structured Interview Framework

1. Tell me about your overall experience in this course? Positive, negative?
2. This course made extensive use of small group activities. Tell me about your impressions of the small group activities.
3. Was everyone in your group more or less equal coming into the course in terms of knowledge and experience?
    a. If no: How did the differences in knowledge or experience affect your small group experience? Did the members of the group with more experience spend a lot of time helping members with less experience? Did that slow down the group, or was it helpful to everyone? Did you find yourself trying to identify where other group members were struggling? [experienced members] Did you find that more experienced group members would go out of their way to help you? [inexperienced members]
4. Would you have preferred to have been in a group that [had everyone at the same level, or had a mix of experience and knowledge]?
5. How could the small group interactions been more effective?

Appendix B: Course Evaluation Form

Class name: Starting Your Career
Class code: 40075

Course Participant Survey

Your name: _____

Please share feedback about instructor 1: <name>	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
This instructor treated participants with respect.	☐	☐	☐	☐	☐
This instructor was easy to hear and understand.	☐	☐	☐	☐	☐
This instructor kept me engaged and active.	☐	☐	☐	☐	☐
Overall, I would take another class from this instructor.	☐	☐	☐	☐	☐

Comments about this instructor:

Please share feedback about instructor 2: <name>	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
This instructor treated participants with respect.	☐	☐	☐	☐	☐
This instructor was easy to hear and understand.	☐	☐	☐	☐	☐
This instructor kept me engaged and active.	☐	☐	☐	☐	☐
Overall, I would take another class from this instructor.	☐	☐	☐	☐	☐

Comments about this instructor:

Next page →

Class name: Starting Your Career
Class code: 40075

Was this course a good fit for your knowledge and skills?
- ☐ Too basic
- ☐ Right for my level
- ☐ Too advanced

How was the length of this course?
- ☐ Too short
- ☐ About right
- ☐ Too long

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Our classroom was a good environment for learning.	☐	☐	☐	☐	☐
The facilities overall (hotel, food, etc.) were good quality.	☐	☐	☐	☐	☐
The materials (slide deck, participant guide, etc.) were clear, error-free, and helped me meet the learning objectives.	☐	☐	☐	☐	☐
This course included a sufficient number of realistic, meaningful real world examples.	☐	☐	☐	☐	☐
This course addressed all of its stated learning objectives.	☐	☐	☐	☐	☐
I learned skills and knowledge in this course that I will use on the job.	☐	☐	☐	☐	☐
Overall, I was satisfied with this course.	☐	☐	☐	☐	☐

What follow-up training (formal or informal) or resources would help you better apply what you learned in this course?

How can we improve this course?

Any other comments you'd like to share about the course content:

Next page →

As you know, there was research going on in this course around small group interactions. We'd like to know a little bit more about your small group experiences.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The small group activities were effective in helping me learn.	☐	☐	☐	☐	☐
When I struggled, I felt comfortable asking my group for help.	☐	☐	☐	☐	☐
I felt like my group supported me when I struggled.	☐	☐	☐	☐	☐
I felt like I did a good job of helping others when they struggled.	☐	☐	☐	☐	☐
I asked questions in my small group.	☐	☐	☐	☐	☐
When I asked questions, I got detailed answers.	☐	☐	☐	☐	☐
When others asked me questions, I gave detailed answers.	☐	☐	☐	☐	☐
Overall, the small group activities were rewarding.	☐	☐	☐	☐	☐

This class naturally features a wide range of incoming experience levels, from no experience to multiple audit engagements. Given your choice, what kind of small group would you prefer to be in?

☐ One made up of participants with an experience level similar to mine.
☐ One that includes participants with a wide range of experience.
☐ No opinion.

Any other feedback you have about the small group activities:

One last thing—you are now starting down the path of a new career with M&P. We'd like to take a quick pulse about how our new associates feel about our Firm.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Overall, I am satisfied with my organization as an employer.	☐	☐	☐	☐	☐
I am proud to work for my organization.	☐	☐	☐	☐	☐
I am committed to doing what is required to help my organization succeed.	☐	☐	☐	☐	☐
I would recommend my organization to my friends and colleagues as a great place to work.	☐	☐	☐	☐	☐