



**Research Report 1965**

**Defining Tailored Training Approaches for  
Army Institutional Training**

**Peter S. Schaefer and Jean L. Dyer**  
U.S. Army Research Institute

**March 2013**

**United States Army Research Institute  
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited

**U.S. Army Research Institute  
for the Behavioral and Social Sciences**

**Department of the Army  
Deputy Chief of Staff, G1**

**Authorized and approved for distribution:**



**MICHELLE SAMS, Ph.D.  
Director**

---

Technical Review by

Thomas R. Graves, U.S. Army Research Institute  
Brian T. Crabb, U.S. Army Research Institute

**NOTICES**

**DISTRIBUTION:** Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: Department of the Army, US Army Research Institute, 6000 6<sup>th</sup> Street (Building 1464 / Mail Stop: 5610), Fort Belvoir, VA 220602-5610.

**FINAL DISPOSITION:** Destroy this Research Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

**NOTE:** The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documentation.

## REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) March 2013	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) October 2010 to August 2012	
4. TITLE AND SUBTITLE Defining Tailored Training Approaches for Army Institutional Training		5a. CONTRACT OR GRANT NUMBER N/A	
		5b. PROGRAM ELEMENT NUMBER 633007	
6. AUTHOR(S) Peter Schaefer and Jean L. Dyer		5c. PROJECT NUMBER A792	
		5d. TASK NUMBER 359	
		5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6 <sup>th</sup> Street, (Building 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6 <sup>th</sup> Street, (Building 1464/Mail Stop 5610) Fort Belvoir, VA 22060-5610		10. MONITOR ACRONYM ARI	
		11. MONITOR REPORT NUMBER Research Report 1965	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A: Approved for public release; distribution is unlimited.			
13. SUPPLEMENTARY NOTES     Subject Matter POC: Peter S. Schaefer			
14. ABSTRACT ( <i>Maximum 200 words</i> ): This report had three purposes. The first purpose was to summarize the research literature and determine the major areas of tailored training research. Those areas were ability grouping, learning in small groups, tutoring, microadaptation, learning styles, and aptitude-treatment interactions (ATI). The second purpose was to determine what types of tailored training were most effective and under what conditions. Of the six areas, only learning styles was deemed ineffective. Each of the remaining areas demonstrated significant tailored training effects. The third purpose was to provide suggestions for tailored training research with near-term applicability in Army settings. Meeting the third purpose required discussing differences between Army institutional training settings and academic research settings which might affect generalizability of the academic research findings. Suggestions for near-term applicable tailored training included focusing on small groups, microadaptation in one-on-one remedial training settings, and ATI research. In ATI, the critical aptitude is prior knowledge. Emphasis is placed on first experimentally assessing the extent and nature of ATI in Army settings and then verifying those findings in classroom settings.			
15. SUBJECT TERMS Tailored Training; Aptitude-Treatment-Interactions; Aptitude Institutional Training; Small Groups			
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	20. NUMBER OF PAGES 77
			21. RESPONSIBLE PERSON Dorothy Young 703.545.2316



**Research Report 1965**

**Defining Tailored Training Approaches for  
Army Institutional Training**

**Peter S. Schaefer and Jean L. Dyer**  
U.S. Army Research Institute

**Fort Benning Research Unit**  
Scott E. Graham, Chief

**U.S. Army Research Institute for the Behavioral and Social Sciences**  
**6000 6<sup>th</sup> Street, Bldg. 1464**  
**Fort Belvoir, Virginia 22060**

**March 2013**

---

Approved for public release; distribution is unlimited.

## ACKNOWLEDGMENT

---

The authors would like to express their gratitude to all of their colleagues who spent so much time listening to and helping refine their concepts.

# DEFINING TAILORED TRAINING APPROACHES FOR ARMY INSTITUTIONAL TRAINING

## EXECUTIVE SUMMARY

---

### Research Requirement:

The Army places a premium upon effective and efficient training. However, what constitutes effective and/or efficient training varies from group to group and individual to individual. For decades researchers have explored the extent to which training quality can be improved by tailoring training, defined as assessing salient individual differences and assigning learners to learning conditions based on those differences. Feasible tailored training research in Army contexts, however, requires an understanding of the academic research in tailored training, a grasp of which methods of tailoring are (in)effective and under what conditions, and an understanding of how differences between Army institutional training and academic research settings and populations might impact generalizability of results.

### Procedure:

A broad review of the academic research literature on tailored training was conducted to isolate the major areas of research. The information was sifted and analyzed to determine which types of tailored training seem to be most effective and under what conditions. The resulting determinations were combined with knowledge of Army settings, resources, and missions to provide suggestions for future tailored training research with near-term applicability in Army settings.

### Findings:

The literature review isolated six broad areas of tailored training research, including ability grouping, learning in small groups, tutoring, microadaptation, learning styles, and aptitude-treatment interactions (ATI). Of the six areas, only learning styles were deemed ineffective. Recommendations for future tailored training research with near-term applicability in Army settings included focusing on learning in small groups, tutoring/microadaptation in one-on-one remedial settings, and ATI research. Specific recommendations regarding ATI research included using prior knowledge as the primary aptitude of interest, and a proposed approach for determining specific ATI in a sampling of Army settings and populations before verifying the feasibility of those findings in specific classroom settings. This 'basic' to 'applied' progression would first be used in more technical subject areas (both cognitive and hands on) which are most similar to the kind examined in the research literature before the 'basic to applied' cycle would be repeated for critical but under-researched areas like decision making and strategic planning.

### Utilization and Dissemination of Findings:

These findings will be disseminated within ARI Benning and to potential research sponsors. Findings will inform best practices for tailored training research.



DEFINING TAILORED TRAINING APPROACHES FOR ARMY  
INSTITUTIONAL TRAINING

CONTENTS

---

	Page
INTRODUCTION .....	1
Rationale for Tailored Training .....	1
Terminology.....	1
Structure of the Paper.....	2
LITERATURE REVIEW .....	2
Ability Grouping.....	2
Learning in Small Groups.....	6
Tutoring.....	10
Microadaptation .....	16
Learning Styles .....	17
Aptitude-Treatment Interactions.....	18
Scaffolding.....	40
LITERATURE REVIEW SUMMARY.....	42
Ability Grouping.....	42
Learning in Small Groups.....	42
Tutoring.....	43
Microadaptation .....	43
Learning Styles .....	43
Aptitude-Treatment Interactions.....	43
ISSUES IN APPLYING RESEARCH FINDINGS.....	44
Subject Matter Domains .....	45
Measuring Prior Knowledge.....	45
Instructor Preparation.....	48
RECOMMENDATIONS FOR FUTURE RESEARCH.....	49
Small Groups .....	49
Tutoring and Microadaptation .....	49
Aptitude-Treatment Interactions.....	50
CONCLUSION.....	51
REFERENCES .....	53



# Defining Tailored Training Approaches for Army Institutional Training

## Introduction

### Rationale for Tailored Training

Research indicates that individual differences in learners impact performance (Jensen, 1998) and interact with learning condition (McNamara, Kintsch, Songer, & Kintsch, 1996). Therefore, what constitutes optimal training varies from individual to individual and group to group. Further, research shows that taking into account relevant individual differences in learning can improve performance (Bloom, 1984; Kalyuga, Chandler, & Sweller, 1998; Kulik & Kulik, 1992). We will refer to the adaptation of training to critical individual differences in learners as *tailored training*.

The realization that individual differences impact learning is not new. More than four thousand years ago the Chinese used competitive tests to select civil servants, and ancient figures like Socrates, Plato, Aristotle, and Quintilian discussed the need to take individual differences into account when designing instruction (Corno, Cronbach, Kupermintz, Lohman, Mandinach, & Porteus, 2002, p. 6-11). What is relatively new is the formalized assessment of individual differences (Jensen, 1998), empirical demonstration of relationships between individual differences and performance (Thorndike, 1985), and experimental examination of how those differences interact with learning conditions (Snow, 1992). However, while the use of experimental evidence has become more common, many researchers feel that changes in instructional methods are often made in the absence of such evidence (Good, 1988; Handelsmann, Egert-May, Beichner, Bruns, Change, & DeHaan, 2004; Pressley & Harris, 1990). Given that lapses in military performance can cause loss of life or damage to expensive equipment, the need for evidentially supporting any tailoring of Army training is paramount.

The purposes of this report are to (1) examine the research literature and isolate the major areas of tailored training research, (2) determine which types of tailored training seem to be most effective and under what conditions, and (3) provide suggestions for future tailored training research with near-term applicability in Army settings. This focus on near-term applicability means that we will not be addressing technology intensive approaches to tailored training like Intelligent Tutoring Systems (ITS). For a recent review of ITS, see Durlach and Ray (2011).

Meeting the third purpose is the most difficult, as it requires considering multiple factors. We must consider the apparent effectiveness/efficiency of the tailoring method, the conditions under which it appears to be effective/efficient, the resource cost of implementing the method, and its feasibility given the differences between Army training environments and the largely academic settings in which the research was conducted.

### Terminology

Diverse terms are used within the academic literature in lieu of tailored training, including adaptive teaching (Corno, 1995), individualized instruction (Fletcher, 1992; Hales, 1978; Heathers, 1977), learner-centered instruction (Kalyuga, 2007), aptitude-by-treatment

interactions (Cronbach, 1957; Snow, 1992), and a variation of aptitude-treatment interactions (ATI) called trait, treatment and task interactions (TTTTI, Berliner & Cahen, 1973). There is also an ambiguity in the phrase individual differences. It is quite common in the research literature to use the term aptitudes. However, some researchers use the term to denote cognitive abilities alone (e.g., Gully & Chen, 2010) while others define aptitudes more broadly to include any learning-relevant psychological characteristic (Carroll, 1967; Corno et al., 2002; Cronbach, 1957; Mandinach & Corno, 1985; Rittle-Johnson, Star, & Durkin, 2009; Shute & Zapata-Rivera, 2008; Snow, 1991, 1992). We adopt the broader usage. Finally, we use the term treatment to mean the structural and presentational properties of instructional methods (Jonassen & Grabowski, 1993).

## **Structure of the Paper**

The review is structured as follows. In the next section, we provide an overview of several areas of academic tailored training research, making note of those methods which seem to be most effective and under what conditions. In the third section, we summarize the major findings of the literature review. In the fourth section, we discuss issues in applying research findings to Army institutional training settings. In the fifth section we outline our recommendations regarding near-term applicable tailored training research. In the final section, we draw together some threads from the paper including considerations of when and why to implement tailored training.

Before proceeding, we make several points regarding the scope of the review. The review provides high-level summaries of different (not all) tailored training approaches with selected illustrations of specific research efforts. The approaches were selected because of their potential relevance to military instructional settings. Research with some tailored training approaches has been conducted primarily in the classroom (e.g., ability grouping) and has focused on specific domains such as reading and mathematics. Other research has been conducted primarily in experimental, non-classroom settings, and with domain-independent/ tasks and/or what might be considered tasks selected for convenience of the experimental design. This review examines some, but not all, of the areas covered in prior general reviews of tailoring and adapting instruction (e.g., Berliner & Cahen, 1973; Cronbach & Snow, 1977; Jonassen & Grabowski, 1993; Wang & Walberg, 1985), and includes some additional major topics, most notably that of human tutoring as a means of tailoring.

## **Literature Review**

### **Ability Grouping**

In classroom settings, attempts are often undertaken to make student groups more homogeneous in terms of ability (Ireson & Hallam, 1999). Known as ability grouping, these attempts are related to tailored training because in some forms of ability grouping, tailoring of instructional methods and materials is involved. This is not always so, however. Ability grouping takes place for a variety of reasons, only some of which are directly related to tailored

training. To gain a feel of the rationale underlying ability grouping, we begin with a consideration of commonly invoked ability grouping arguments, both pro and con.

**Pro-and con-ability grouping arguments.** Turney (1931, as cited in Slavin, 1990; see also Ansalone, 2009) lists four con- and seven pro-ability grouping arguments. Critics of ability grouping posit that grouping impedes the progress of slow pupils by depriving them of the stimulating presence of more able students. Further, it is possible that ability grouping stigmatizes membership in the lower ranking groups, thus discouraging those pupils. In addition, teachers object to being assigned to the lower ranking groups, which may further depress academic growth. Finally, critics say that ability grouping is irrelevant, as teachers lack the time or other resources to tailor material for groups of different ability levels.

Supporters of ability grouping argue that grouping can facilitate tailoring by allowing teachers to utilize teaching methods which are appropriate to the group as a whole. Grouping may also enable individualized instruction when dealing with smaller groups of slower students. Grouping may help maintain interest and incentive among bright students. Grouping might encourage slower students to participate more. Ability grouping plausibly helps students of all ability levels to progress commensurate with their abilities. Grouping may reduce failures. Finally, grouping is thought to simplify the task of teaching by allowing the teacher to focus his or her attention upon the needs of students who are more homogeneous in relevant characteristics.

**Causal hypotheses in ability grouping.** Within the ability grouping literature, there is a general consensus that ability grouping can impact performance via several paths. Three different kinds of effects have been proposed: social, institutional, and instructional (Pallas, Entwisle, Alexander, & Stluka, 1994).

Social effects involve the impact that peers (other members of an ability group) have upon perceived academic norms and self-expectations of a student (Pallas et al., 1994). Institutional effects involve the impact that knowledge of a student's ability group placement can have upon teacher expectations and behavior. Instructional effects involve the impact of quantity, quality, and pace of instruction on learning (Gamoran, 1986; Oakes, 1985). Evidence for social effects was found in first grade reading groups (Eder, 1981; Femlee & Eder, 1983). Inattentiveness and disruptive behavior was common in the lower ability reading groups, taking away valuable teacher time from instruction and toward disciplinary actions. Brophy and Good (1970) found indirect evidence for institutional effects in that teachers interacting with students perceived as high achieving demanded and praised better performance more often than teachers interacting with students perceived as low-achieving. Although this effect is not always found (Alexander & Cook, 1982; Carbonaro, 2005; Weinstein, 1976), this does raise concerns about teacher behaviors impacting student performance (Chorzempa & Graham, 2006). However, evidence indicates that instructional variables are the main factors behind performance differences (Brody & Mills, 2005; Gamoran, 1986; Hallam, 2002; Ireson & Hallam, 1999; Pallas et al., 1994; Whitburn, 2001).

**Methodological issues in ability grouping research.** Despite the amount of research focused upon the effects of ability grouping, there are significant methodological problems to be

overcome—problems which are fully acknowledged by researchers in this field. Ireson and Hallam (1999) note multiple problems in coming to conclusions regarding the effects of ability grouping. For example, it is often difficult to categorize a given school as using one form of grouping over another, because multiple grouping approaches are sometimes used within the same school. Further, research suggests that effects may not be the same in different domains or between teachers or across time. In addition, there may be complex interactions among grouping approaches, teaching methods, and teacher attitudes. Finally, large-scale changes in education as a whole may be driving apparent ability group effects.

Lou et al. (1996, as cited in Lou, Abrami, & Spence, 2000) further note that ability grouping studies often vary in outcome measures (e.g., standardized tests or teacher-developed), methodology, intensity, duration, and grade level. In addition, students can be assigned to groups on the basis of perceived academic ability (Gamoran, 1986), general ability measures (Slavin, 1987) or subject specific measures (Slavin, 1990). Finally, there are different types of ability grouping approaches (see below) which vary in their effectiveness. In sum, claiming success on behalf of ability grouping is difficult (Hallam, 2002).

**The effects of ability grouping.** The generalizations we draw are tentative. As noted above (Hallam, 2002; Ireson & Hallam, 1999), there are a variety of factors which appear to moderate the effectiveness of ability grouping. For example, take the thesis that there may be complex interactions among grouping approaches and teaching methods. A reasonable way of parsing such data would be a combination of meta-analytic methods and identification of moderator variables to establish and then qualify general trends. Unfortunately, this has not been done. On the one hand, there are several meta-analyses of ability grouping studies which provide some guidance regarding effectiveness. On the other hand, these meta-analyses are qualified to an unknown degree because the moderating variables are not well understood. To give the reader a birds-eye view of these findings, we rely most heavily on meta-analyses, but also include specific studies which appear to contradict the meta-analytic conclusions.

**Types of ability grouping.** The following classification scheme for ability grouping is drawn from the meta-analysis of Kulik and Kulik (1992). In each case, the effectiveness of a given approach was demonstrated by performance differences between the ability-grouped students and comparable students in mixed-ability (non ability-grouped) classes. As the selection criteria for entry into ability groups differed (as noted above), the measures upon which the comparability of students was determined also differed. Effect size estimates were derived by dividing the mean differences between the groups on some criterion by the pooled standard deviation of performance on that criterion. The meta-analytic results cited below were conducted with both primary and secondary school students. While not explicitly stated, the majority of the domains examined have been either mathematics or reading. In all cases, standardized measures of reading, math, etc. were used to estimate effect sizes.

***Multi-level class approach.*** In this approach to ability grouping, students in a given grade are selected on the basis of either a subject-specific standardized test or a general ability measure and placed in groups. These groups are then taught in separate classrooms, for either all subjects or for a single subject. For example, all high, average, and low performing seventh

grade readers might receive English instruction in separate classrooms for the first class of the day.

Effect size estimates for the multilevel class approach center around zero for both reading and mathematics (Kulik & Kulik, 1982, 1992; Slavin, 1987, 1990, 1993). However, other studies have found lower ability students are hindered by ability grouping (William & Bartholomew, 2004; Ansalone, 2001), although this is sometimes localized to specific subjects (e.g., mathematics but not general science or English: see Ireson, Hallam, Hack, Clark, & Plewis, 2002). There thus appears to be a discrepancy in the literature regarding the multilevel class approach. In addition, Kulik and Kulik (1992) note that the multi-level classroom approach often did not involve tailoring of instructional materials or methods, but rather was used to reduce student variation in classes. In fact, in some of the older studies, teachers were instructed to keep the content constant across ability groups. In more recent studies, content adjustment was informal and up to the discretion of the instructor. Given these confounds, we remain agnostic on the effectiveness of this approach.

***Cross-grade approach.*** In this approach, students from several grades are grouped together on the basis of achievement. They are then taught in separate classrooms without regard to regular grade placement. For example, a high-performing seventh grader might be taught in the same group as a low-performing tenth grader. Here, Kulik and Kulik (1992) found an average effect size of .30. They note that this approach uses some degree of differentiated instruction in that the materials vary from group to group. For example, high performing third graders might be using materials suitable for an average seventh grader, compared to average or below average third graders using third grade or first grade materials, respectively. Some caution is due here. Most of the cross-grade grouping studies focus on the Joplin Plan, a cross-grade grouping approach to reading in the elementary grades. Therefore, the effect size estimate may vary with other subject matter and grades.

***Within-class grouping approach.*** This is perhaps the most common grouping method (Lou, Abrami, & Spence, 2000). Here, a teacher forms ability groups within a single classroom and then attempts to provide each group with instruction appropriate to group aptitude. As Kulik and Kulik (1992) note, this implies differentiated instruction of some sort. Otherwise, why divide the classroom into 3 groups and then give the same presentation to each group? Kulik and Kulik (1992) found an average effect size of .25.

***Enriched classes approach.*** In these classes, the students receive more varied educational experiences than would be available to them in the regular classroom curriculum for their age. Kulik and Kulik (1992) found an average effect size of .41.

***Accelerated classes approach.*** Here, students are given more advanced materials. But this approach also allows the students to proceed more rapidly or finish schooling at an earlier age. These studies yielded the largest average effect size (.87), but this only held true when comparing accelerated students of a given age with same-age controls. When comparing them to older, otherwise comparable students, there was essentially no difference (mean effect size of .02).

**Research implications.** There are various problems in assigning cause-effect relationships to ability grouping. However, a few tentative conclusions can be drawn. First, simply grouping people of similar ability levels together without further tailoring of materials or method does not appear to impact performance.

Second, the more tailoring that takes place, the better performance will be (Eash, 1961; Jones, 1948; Kulik & Kulik, 1982, 1992). This is not surprising, as those ability grouping approaches which involve the most tailoring (accelerated and enriched classes) also tend to have smaller class sizes (Sausner, 2005). These methods are arguably most similar to individualized tutoring, which is seen as the most effective tailored training approach (Bloom, 1984).

Third, within-classroom grouping seems to be the most common method although it is not the most effective. No doubt there are resource constraints at play here. In the multilevel class and cross-grade grouping approaches, significant demands are placed upon schools in terms of complex scheduling and multiple classrooms. Approaches at the other end of the spectrum—enriched and accelerated classes—tend to receive little funding (Sausner, 2005) and involve significant involvement from the teachers and commitment in terms of tailoring materials. Given that within-classroom grouping is effective and is less resource intensive, greater research into effective ways of utilizing within-classroom grouping in Army settings is warranted.

### **Learning in Small Groups**

The research on learning in small groups differs from the research on ability grouping in that small groups are not necessarily distinguished by ability and ability groups are often larger, sometimes comprising an entire class (i.e., twenty-five or more individuals). In this review, a small group means less than ten members, more typically four to five. As with ability grouping, small group learning can take more than one form (e.g., students working jointly to solve science or math problems, which have well-defined solutions, or working jointly to develop a complex plan or strategy, where the solution is not well-defined). Most research has been conducted with more routine tasks. The research includes a variety of small group learning modes where the same concept can have different names and the same name can have different meanings.

Within the public school system, the phrase “cooperative learning” is applied to small group methods developed to improve achievement and to enhance social skills (Bossert, 1989; Cohen, 1994; Johnson & Johnson, 2009; Johnson, Maruyana, Johnson & Nelson, 1991; Slavin, 1980, 1991). The link to tailored training is that “cooperative learning” is viewed as one means of accommodating individual differences in classrooms, particularly students who are struggling with the group task (Antil, Jenkins, Wayne, & Vadasy, 1998; Cohen, 1994). This concept became accepted in the 1980s (Johnson & Johnson, 2009), and typically involved distinctions among three reward structures for individuals within a group (Bossert, 1989; Johnson et al. 1991): cooperative (based on quality of group work), individualistic (based on each individual’s work), and competitive (intragroup dependency among rewards with high rewards for one individual meaning low rewards for another).

However, learning in small groups also means that teachers delegate some of their authority. Therefore they have less control over the tailored training that will occur. Certainly tailoring to individual differences in the sense of addressing students' weaknesses or enhancing students' strengths is not guaranteed in small group learning settings. In fact, in an early commentary on cooperative learning, Berliner (1985) viewed this research as focused on finding the best approach, which was typically a cooperative reward structure (Bossert, 1989), not on systematically determining which group methods work best for different kinds of students or how tailoring can be achieved in these settings. This comment by Berliner applies to later research as well.

For other researchers, "cooperative" is defined in terms of the division of labor within a group, with distinctions drawn between cooperative and collaborative groups (Dillenbourg, Baker, Blaye, & O'Malley, 1996; Shute, Lajoie, & Gluck, 2000). With cooperative groups, each individual works on a separate part of a group task and then merges the effort with others to complete the task. With collaborative groups, there is mutual sharing, negotiation, and engagement among group members to complete the task. Also collaboration settings should be highly interactive, with members influencing each other during the problem-solving process (Dillenbourg, 1999). Collaboration is typically associated with synchronous communication, while cooperation is often associated with asynchronous communication. Lastly, Shute et al. (2000) identified competitive groups, referring to situations where groups compete with each other. This is intergroup competition, not the intragroup group reward structure referenced in the cooperative learning literature. Again, how these variations directly impact tailoring has not been specified.

**Group member interactions and performance.** We lack a systematic body of research on how tailoring can be achieved within a small group structure. As stressed by Bossert (1989) and Cohen (1994), research on group dynamics and the sequence of group member interactions is needed to determine how small groups work, why they are successful, and to explain inconsistencies in more general research findings. However, there is some research on the dynamics of group member interactions and on the impact of some student characteristics on learning in small groups which provides insight into the issues associated with tailoring training in small group settings. We provide examples of such research next.

Although small groups in public school classrooms are typically not based on ability, the effect of different combinations of ability on achievement with well-defined tasks has been examined (Shute et al., 2000; Webb, 1991; Webb, Nemer, Chizhik, & Sugrue, 1998). In general, low-ability students perform better in heterogeneous groups; with high- and low-ability students often forming a teacher-learner relationship. Medium-ability students work best in a homogeneous group. The results for high-ability students are not as consistent, with research showing high achievement in both homogeneous and heterogeneous groups.

Webb conducted multiple efforts specifically examining the role of explanations within small groups (e.g., Webb, 1982, 1984, 1991). Individuals in these groups (typically junior high school students) worked together to solve math problems and were told to not divide their work but to help anyone in the group who had difficulties. A teacher was available only when the group needed help. Often group composition reflected different mixes of ability levels to

determine if these mixes impacted achievement and/or group interaction. Typically, there was no comparison group with the teacher as the primary instructor. Webb found that individuals who gave explanations to other group members achieved more than those who did not offer explanations. In contrast, individuals who received nonresponsive feedback from others, i.e., no explanation in response to an error or a question and/or received the correct answer without an explanation, learned less than individuals who received responsive feedback. In summary, giving explanations to other group members was positively related to achievement and receiving nonresponsive feedback from group members was negatively related to achievement.

Some research on student collaboration/cooperation has been conducted in conjunction with tutoring (Chi, 2009; Craig, Chi & VanLehn, 2009). In Chi's research, the comparison involved pairs of students watching a tutoring session and simultaneously collaborating to solve a problem versus a single student watching the same tutoring session or a worked example, and/or pairs of students collaboratively observing worked examples. Results showed that the "active observing" process was most effective when pairs (not a single individual) solved problems while observing tutoring and the videos involved high-ability tutees. With Craig et al. (2009), the comparisons were an individual observing a worked example, pairs who collaboratively observed worked examples, and pairs who collaboratively observed tutoring. The most effective group, based on long-term retention measures, was the collaboratively-observing-tutoring condition.

Perceived status of group members is known to impact group interactions (Cohen, 1994). It is perceived, not actual, status that counts, with those having a higher perceived status more likely to be perceived as the leaders. They are likely to become more active in the group process compared to others. Inequalities in participation are, in turn, linked to differences in achievement gain.

The task or problem mostly likely impacts group interactions as well (Cohen, 1994). Group tasks that involve ill-structured problems, where no single individual can solve the problem (Shute et al., 2000), probably create different patterns of interactions than group tasks which are "routine" and can be solved by a single individual (Cohen). However, we did not find any comparative studies examining potential differences in member interactions with well-defined versus ill-defined tasks.

**Research implications.** In the Dyer, Wampler and Blankenbeckler (2011) investigation of tailored training in Army functional and professional courses, use of small groups was cited very frequently by instructors as a means of addressing individual differences and tailoring instruction. The instructors probably viewed small groups as a macro-means of tailoring as it differs from the one-size-fits-all lecture approach. A caveat is that the information was obtained from instructor interviews, as there was no opportunity to directly observe the small group settings within military classrooms. Consequently, details on the extent and nature of tailoring were not available. The group tasks and the ways groups were formed differed substantially from that in the research literature, an apparent consequence of the different student population and the instructional goals. A summary of the findings is presented next.

***Composition of small groups in military settings.*** Instructors cited different means of determining group composition. The most common method was to ensure each group had a highly experienced/skilled student in the subject area, which allowed the more experienced/skilled person to provide peer-to-peer assistance. Another technique was to organize the students so disparate experience/skills (e.g., heavy force/light force) were uniformly distributed throughout the groups. Some groups were organized based on the anticipated requirements of the student upon course completion. For example, when students with varied ranks (e.g., junior enlisted to field grade officer) attended a single course, the groups could be organized so each rank group was represented, as would be the case in a military unit. In this regard, Cohen's (1994) discussion of the impact of perceived status on group processes is relevant and worthy of investigation in military instructional settings. When students were of similar ranks, the objective of some groups was to have students participate in exercises that placed them in positions similar to those they could hold after graduation. Here, grouping required students to perform as a team or staff member or for individual students to accomplish tasks, perform operations, conduct planning, or make estimates as they would in duties after course graduation. Duties could be rotated among students from exercise to exercise, permitting students with the opportunity to practice aspects of their primary responsibilities, as well as gain first-hand knowledge in related tasks and skills through supporting roles. It should be noted that in some instances, assignment to a group was merely by convenience with no systematic means used.

***Training objectives and group processes in military settings.*** One objective of the small group instruction was to have students work on multiple-solution type tasks such as producing a complex plan where there was more than one feasible solution (see Dyer et al., 2011). These tasks correspond to Cohen's (1994) category of "true-group" tasks.

Of considerable interest is that a few instructors (Dyer et al., 2011) implemented grouping for the multiple purposes of cooperation, collaboration, and competition, even within the same phase of training; groupings consistent with the concepts cited by Dillenbourg et al. (1996) and Shute et al. (2000). In one situation, students were first assigned to groups where they were required to cooperate to complete a complex task or problem. Each member had a designated part of the larger task to complete individually. For example, one person might develop the plumbing requirements for a facility, while another determined the electrical requirements, while another examined the structural feasibility for the location and intended purpose of facility. Once the individual parts were completed and assessed by an instructor, the group collaborated by combining the various parts and sharing ideas. They worked together as a team to modify the final product plan so that it represented the best ideas from each part and all parts melded into a single integrated product that met all functional requirements. Finally, each group presented its product plan for review and questions from other groups and the instructor. Each group was scored on the overall product plan, according to criteria specified by the instructor. The group with the "best" product plan according to the scoring criteria would win the competition and receive recognition.

***Instructor preparation.*** One focus of the cooperative group research in the public schools was on how to prepare teachers to ensure that cooperation and collaboration occurred within a group yet have each student be accountable for his/her performance. However, none of

the military instructors interviewed by Dyer et al. (2011) indicated difficulty in getting students to cooperate or collaborate. In addition, instructors had no difficulties in systematically assigning individuals to groups. Yet if tailored training is to exist within military small-group settings, by adapting to critical individual differences during the actual instructional process or by making groups function in ways that specifically address individual differences, then instructors must be trained on how to facilitate such tailoring. However, we need dedicated research on these instructional techniques rather than relying on a teacher's creativity to come up with a solution (Berliner, 1985).

***Generalization of research findings.*** Clearly, major distinctions exist between the training settings investigated in the research on small-group learning in school systems and small-group instructional settings in the military (e.g., group tasks, training objectives, and group composition). As there is minimal research on small-group instructional processes within the military similar to that conducted by Webb (1982, 1984, 1991) and Chi, Roy, and Hausman (2008), the extent to which these findings generalize to military instruction is not known. Given the frequent use of small groups in military training in conjunction with such uncertainties regarding what makes small-group learning effective in the military and how these settings can be used for tailoring training, further research appears to be a fruitful area of investigation.

## **Tutoring**

**Human tutoring.** Tutoring, defined as “teaching or guiding, usually individually, in a special subject for a particular purpose” (Merriam-Webster, Inc., 2002) with a tutor being a private teacher, could be conceived as the best mode of tailored training. In general, tutoring specifically adapts to the status and needs of each individual or two to three individuals. It differs from microadaptation (described in the next section) in that tutoring typically involves planned sessions with an assigned individual, a tutor, rather than being a direct, immediate response by a classroom teacher to a student's questions, actions, or behavior.

***Effectiveness of tutoring.*** Bloom's (1984) classic article stressed the advantages of tutoring and how tutoring could raise the level of achievement by two standard deviations, although the thrust of the article was on determining less costly means of instruction that would achieve the same objective (i.e., also known as learning for mastery [Anderson, 1985]). In a meta-analysis of mastery learning programs, Kulik, Kulik, and Bangert-Drowns (1990) found sizeable effects with the mastery programs (effect sizes of 0.50 to 0.80) which had special characteristics such as high mastery standards, more quiz feedback, and group-based settings, but none achieved the substantial effects which Bloom cited with tutoring. In a recent review of tutoring, VanLehn (2011) also stated that the effect size was lower than Bloom reported, approximately 0.79, with tutoring being more effective than standard instruction without tutoring. In summary, the research consistently shows that tutoring is effective (Cohen, Kulik, & Kulik, 1982; Elbaum, Vaughn, Hughes, & Moody, 2000; Graesser, Sidney & Cade, 2009; Shanahan, 1998), even with tutors (older students, paraprofessionals, and/or adult volunteers) who have not been trained extensively (Graesser & Person, 1994; Person & Grasser, 2003, Ritter, Barnett, Denny, & Albin, 2009).

Certain tutoring conditions have produced effect sizes greater than 0.40. Structured programs were more effective (Cohen et al., 1982; Ritter, Barnett, Denney, & Albin, 2009; Shanahan, 1998). With structured programs, tutors had specific lessons and materials to cover compared to unstructured programs where tutors had minimal training and tutors and students often simply read together (Ritter et al., 2009). Cohen et al. (1982) cited more gains in mathematics than reading. Tutors who had training were more effective (Cohen et al., 1982; Elbaum et al., 2000; Ritter et al., 2009; Shanahan, 1998).

Tutoring elementary school children in reading or mathematics in public school setting has often been the research focus (e.g., Juel, 1996; Lepper & Woolverton, 2002; Ritter et al., 2009; Shanahan, 1998), as opposed to tutors teaching new material or skills. Nonetheless there is a body of research on tutoring new material with older students. Some examples of this tutoring research are: computer programming with college students (Merrill, Reiser, Merrill & Landes, 1995); research methods with college students and algebra with seventh graders (Graesser & Person, 1994); circulatory system with eighth and ninth graders (Chi, Siler & Jeong, 2004); physics with college students (Siler & VanLehn, 2003); and mathematics with ninth and tenth graders (McArthur, Stasz, & Zmuidzinas, 1990).

***The tutoring process.*** Shanahan (1998) posed the central question of why tutoring works. It appears there is relatively little comparative information on what conditions make tutoring effective, as controlled experiments in tutoring are extremely difficult to execute (Graesser et al., 2009). In addition, there is much less research on the tutor-student process compared to the instructional process used by classroom teachers. Yet extensive, detailed examinations of tutor-student interactions/dialogue based on video or audio tapes of tutoring sessions have been conducted. The accumulated findings from this research are providing insights into tutoring processes. Of interest is that much of this research has been stimulated by intelligent tutoring systems (ITS), whose development requires an understanding of the tutoring process (Lepper & Woolverton, 2002; Person & Graesser, 2003; VanLehn, 2011). One of the goals of ITS developers is to build a system that is as good as a human tutor. However, typically, computer-based systems have been shown to be less effective than human tutors (VanLehn, 2011). Consequently, researchers have found it important to know what a good tutor does and what happens during the tutoring process.

Researchers have established specific coding schemes to document the dialogue between tutors and students (e.g., Chi, Siler, Jeong, Hamauchi & Hausmann, 2001; Person & Graesser, 2003; Putnam, 1987). There is no commonly accepted coding scheme, although typically percentages of time tutors and students talk, whether tutors provide explanations, ask questions, provide feedback, etc. are obtained. Schemes vary with respect to the codes used to depict what tutors and students actually say (e.g., depth of scaffolding, types of examples the tutor provides, difficulty of exercises, type of feedback, whether the tutor focuses on the student's understanding of content versus just getting the correct answer, whether the student spontaneously explains content, types of student questions). Coding schemes can also be tailored to the subject matter. Summarized below are conclusions regarding the tutoring process by major researchers in the field who conducted extensive analyses of tutor-student dialogues.

Person and Graesser (2003) cited 14 findings about human tutoring based on 15 years of research. A caveat about their findings is that they explicitly stated that the tutors they studied were “not trained to use sophisticated tutoring techniques, but rather were representative of tutors that generally do most of the tutoring in school systems” (p. 1). Major findings included: tutors do not typically use Socratic strategies; tutors often provide hints and elaborations; feedback is typically immediate and sometimes the feedback is not responsive to the student’s error; tutors rarely attribute errors to lack of ability; compared to typical classroom settings, tutors and students ask more questions; and good students ask better, not more, questions. Consistent with these conclusions are findings by Graesser et al. (2009) that despite tutors receiving some training, tutors are not likely to implement advanced tutoring strategies such as Socratic reasoning, building on prerequisite knowledge, scaffolding techniques, diagnosis, or asking why, how, and what if. Typical exchanges between the student and tutor can consist of short answers and responses (Graesser et al., 2009; Graesser & Person, 1994; Putnam, 1987) with the goal of getting the student to answer correctly without necessarily determining the reasons for the student’s answers or questions. Chi and associates (Chi & Roy, 2010; Chi et al., 2001; 2004) also refer to the tendency of tutors to be didactic, i.e., to give many explanations which may not be directly responsive to students’ needs, and to be unable to take the perspective of the student. These researchers concluded that tutors are not optimally adaptive when these conditions exist in the tutoring process.

In VanLehn’s (2011) review, eight hypotheses were offered regarding what makes tutoring more effective compared to classroom instruction or computer tutors. VanLehn rejected six of these hypotheses based on the research reviewed. These hypotheses and VanLehn’s corresponding conclusions were as follows:

- Human tutors are able to diagnosis the student’s misconceptions and then adapt their tutoring accordingly. VanLehn concluded that tutors do identify student errors, but typically do not determine the reason for such errors.
- Compared to computer tutors, human tutors are more skilled in selecting the types of tasks, including task difficulty, appropriate for each student vice following a more or less prescribed curriculum. However, tutors often follow a “script” versus selecting tasks appropriate for each student.
- Human tutors use sophisticated instructional strategies such as Socratic reasoning. VanLehn concluded that research examining tutoring dialogues has shown that use of sophisticated instructional strategies is rarely the case.
- In a tutoring session, the student is allowed to take much of the initiative. Although students do take the initiative more frequently than in a classroom environment, it is not necessarily at a high level.
- Human tutors have a broader and deeper understanding of the domain. Although VanLehn concluded this may be the case, often human tutors do not offer deeper explanations, unless cognitive skills are taught.
- Human tutors are able to motivate students. Even though tutors praise students, VanLehn indicated that computer-mediated text seems to be as effective, in some cases.

The two hypotheses that VanLehn considered as possible reasons for the effectiveness of tutoring were feedback and scaffolding. Feedback allows students to monitor their thinking and make repairs; and the immediate feedback typical of tutoring sessions makes it easier for students to modify their concepts and knowledge. Scaffolding, meaning guided prompting by the tutor that pushes the student to understand the material on his/her own rather than telling the student the answer, has also been shown to be effective and promotes deeper thinking by the student. VanLehn hypothesized that another reason tutors are more effective is due to “granularity,” with granularity referring to the amount of reasoning that can be required or is possible with tutor and student interactions.

Lepper and Woolverton’s (2002) summary of ten years of research on tutoring profiled highly-effective tutors differently than Person and Graesser (2003) and VanLehn (2011). The tutors were specially-selected individuals who tutored elementary students (in first through sixth grade) in mathematics. Based on student performance, the most effective tutors within this group were then compared to the other selected tutors who had similar backgrounds but whose students did not perform as well. A point stressed in this review, yet not cited in the reviews by VanLehn and Person and Graesser, is that the highly-effective tutors focused simultaneously on both cognitive and motivational factors. Lepper (Lepper & Woolverton 2002) coined the acronym INSPIRE to highlight the seven major characteristics of the most effective tutors in their research: intelligent, nurturant, Socratic, progressive, indirect, reflective and encouraging. Summaries of each characteristic are given next.

- **Intelligent:** The most effective tutors showed depth and breadth of knowledge, used very effective examples and models, applied subject-specific pedagogical knowledge, and articulated the reasons for the instructional and motivational techniques they used.
- **Nurturant:** The tutors were very supportive throughout the tutoring sessions, and showed confidence in the student’s ability.
- **Socratic:** The best tutors asked questions rather than providing directions or assertions; they avoided directly giving student answers and persisted in an asking-question strategy which led the student to the correct answer. The best tutors also had a better understanding of the students’ errors and ignored small, non-critical errors, whereas the less successful tutors typically responded to every error. The most effective tutors capitalized on errors when they thought the students would benefit from learning from their mistakes.
- **Progressive:** Tutors deliberately planned sessions with problems of increasing difficulty which were also aimed at diagnosing students’ level of knowledge, as well as problems that could detect students’ misunderstandings.
- **Indirect:** The authors stressed that the tutors systematically avoided overt criticism, yet typically did not enthusiastically praise students.
- **Reflective:** Tutors wanted students to understand basic concepts, and often used the technique of having students reflect out loud on what they had done and explain their answers.
- **Encouraging:** Tutors used a variety of techniques to bolster students’ feelings of competence, to challenge students, to stimulate students’ curiosity, and to give students a sense of control in the tutoring process.

***Tutoring examples.*** The examples presented in this section illustrate how tutoring techniques can vary. In Juel's (1966) research with remedial tutoring of first-graders in reading, two factors distinguished successful tutor-student dyads from less successful dyads: use of scaffolded experiences and cognitive modeling of reading processes. With scaffolded experiences the tutors provided just enough information to enable the students to do tasks on their own. The more successful tutors had more of these experiences than the less successful. With modeling, the tutor elongated the sequence of sounds in words with which the students had difficulties. Juel pointed to some other relevant impacts of tutoring in this particular context. The tutors were student-athletes who had reading problems and the students were from a very low-income area. Juel reported there was affection and bonding regardless of the success of the dyads. This relationship was important for the children, with tutors indicating they could identify with the children due to their own reading problems when they were young. These behavior patterns correspond in many ways to the Lepper and Woolverton (2002) summary.

Wittwer, Nuckles, Landmann, and Renkl (2010) examined the nature of explanations that tutors used. Only when tutors had information about the students' prior knowledge were they able to individualize the instruction and customize the explanations they provided and the questions they asked.

In the Merrill et al. (1995) research, tutoring was with new material, specifically training college students on the LISP programming language. Tutors provided positive confirmatory feedback quickly, a finding consistent with Juel (1966). The analysis focused on the type of student errors and how both students and tutors dealt with them, what the authors called "error repair." Students themselves were typically aware of low-level type errors, but only the tutors identified the more complex errors. Tutors also varied their feedback in accordance with the nature of the error. Explicit and quick corrections were given for errors that did not provide a significant learning opportunity. On the other hand, tutors provided less support for errors that offered significant benefits to a student if the student could solve the problem alone. They would draw the student's attention to a possible area that would solve the problem but not formally tell the student the answer. Thus "as learning consequences increased, the tutors allowed the students to perform more and more of the error recovery" (Merrill et al., 1995, p. 358). In addition, the tutors did not diagnose the cause of student errors. Instead they focused on correcting and reviewing relevant curriculum material; they kept the students on track. Putnam (1987) found a similar emphasis on progressing through the curriculum versus using extensive diagnosis and remediation.

In Graesser and Person's (1994) analysis of tutor-student dialogue, they estimated that student questions were 240 times more frequent during a tutoring session than a typical classroom session. A high proportion of the tutor-student dialogue was short "yes/no" situations. Less frequent were interchanges involving why, what if, and why not, which the authors called deep-reasoning questions. Of interest is that these questions were posed by both students and tutors. Over 60% of the questions were attempts to ensure "common ground" (p. 125). In other words, tutors spent time ensuring what the students understood, while students spent time ensuring they had the correct information and understanding. Establishing common ground is important as Nickerson (1999) documented the most frequent communication error is falsely

assuming others have the same knowledge. Lastly, Graesser and Person categorized 30% of the student questions as a form of self-regulation in that students focused on correcting their own potential comprehension and knowledge errors.

These studies indicate some ways tutors adjust their feedback mechanisms and instruction to the individual student. They support the conclusions reached by Merrill, Reiser, Ranney and Trafton (1992) in their comparison of human and computer tutors. Human tutors assist students by having them do more of the error recovery process whereas computer tutors take on more of the error repair process. Human tutors are more flexible in their level of assistance than computer tutors. The assistance and control from a computer tutor is more noticeable.

Although we may think that the tutor is the key to effective tutoring processes and outcomes, Graesser et al. (2009) stressed that effectiveness also depends on the student's status. For example, students learn more when they contribute ideas to the tutoring session. They stated that students would benefit if they learned how to ask specific questions, questions that better reflect deficits in their knowledge and understanding of the subject matter. Some of Chi's research described below would indicate that what the tutor asks can influence the extent to which students reveal what they know to the tutor.

Chi's (2009) article presented a conceptual framework for what learners do, specifically observable behaviors or activities. Four distinctions were drawn: learners being passive, active, constructive, or interactive. Passive referred to behaviors such as watching or listening. Active referred to doing something physically such as underlining text or taking notes. Constructive behavior meant the student produces outcomes that go beyond the information presented. Interactive behavior requires an extensive dialogue with another on the same topic and considers that individual's contributions. The general thesis was that interactive activities are best for learning, then constructive, then active, and lastly passive. This perspective of looking at what students do was examined in earlier work on tutoring situations by Chi et al. (2001). This research may help explain some of the differences between the conclusions by Graesser and VanLehn regarding the tutoring process and the conclusions by Lepper and Woolverton.

The first phase of the Chi et al. (2001) research examined tutors (college students) working with eighth graders on the circulatory system. The typical pattern of tutors talking more than students, relatively short turns by the tutor and student, and tutors controlling the turns and giving explanations was found. In other words, the process was more didactic than Socratic. Also, the tutors did not attempt to understand student's misconceptions, and there was little constructive activity on part of the students. The second phase of the research was intended to see how tutors could elicit more constructive activity on part of the student. Specifically, tutors (the same tutors as in the first phase) were told to suppress giving explanations, feedback, and extra information. Instead they were given a list of content-free, open-ended prompts to ask students, e.g., "Could you explain this in your own words?" "What do you think?" "What's going on here?" "Could you connect what you just read with what you have read before?" The result was a substantial change in the dialogue between the tutor and student. Under these conditions, students talked more than the tutor, the exchanges were longer, and tutors became more

interactive and less didactic, with more deep scaffolding prompts. Training time and learning outcome were the same.

In summary, the tutor-student information flow and feedback loop changed. Chi et al. (2001) concluded that the students in the prompting condition were more constructive because of the frequent tutor prompts and scaffolding. In turn this allowed the students to display more of what they knew and didn't know. Thus the tutors could more accurately evaluate the student, which then allowed them to pursue extended scaffolding. The students in the non-prompting group (first phase of the research) did learn, but this was attributed to learning from hearing explanations and repetition of information, not because the tutors were more adaptive. If one goal of tutoring is to diverge from what tends to be a didactic process with short tutor-student interchanges, then having tutors use content-free, open-ended prompts is potentially one way to move toward a more Socratic process that engages students more completely and enables tutors to adapt better to the status and needs of the student.

**Research implications.** Three research limitations were cited by Graesser et al. (2009). First, studies with accomplished tutors are uncommon and have small sample sizes. Second, there are limited “detailed analyses of human tutorial dialogue that are related to outcome measures and that have a large sample of tutors” (p.11). Third, there is little research comparing tutors with different levels of expertise, as most of the research compares tutoring to conventional instruction. Another research gap salient for our purposes is the limited to no research on domains with a heavy hands-on component, typical of many military tasks.

Dyer et al. (2011) found that with highly technical military content or domains with a heavy hands-on component, Army instructors indicated that small groups of three-to-four students were formed with an instructor assigned to work with each group –situations which approximate many tutoring situations. In addition, one-on-one instructor-student conditions often occurred with hands-on practical exercises. Instructors stated these procedures allowed them to adapt their instruction to student needs. In this research effort, the teacher-student interactions were not observed. However, given that the instructors typically said they were not trained on how to individualize instruction and that military instructors usually serve in an instructor position for only two to three years, it could be expected that the dialogue is typical of much of the tutoring research. Nonetheless, research is needed to verify if this is the case, as the student population and the instructional goals (i.e., the tasks and skills essential to a profession) differ substantially from those found in the extant research. When designing training that prepares military instructors to tutor, then the content-free, open-ended prompts approach might be successful in enabling them to be adaptive and to better diagnose and correct student misunderstandings.

## **Microadaptation**

Microadaptation has been defined as “continually assessing and learning as one teaches-thought and action intertwined” (Corno, 2008, p. 163). In other words, microadaptation refers to instructor delivered on-the-spot detection and repair of student errors or misunderstandings. References to microadaptation can be found as far back as the days of the Roman Empire. The orator Quintilian (trans. 1920, as cited in Corno, 2008) even expressed ideas very similar to

Vygotsky's (1978) zone of proximal development. Modern approaches to microadaptation often revolve around the idea of accommodation: the two-fold process of capitalizing on student strengths while circumventing student weaknesses. Both anecdotal and first-hand observation of instruction demonstrates that microadaptation is an everyday phenomenon (Corno, 2008; Lampert, 1985; Nuthall, 2004). Because micro-adaptation requires lengthy teaching experience in a given domain as well as extensive content knowledge of that domain (Putnam, 1987), it is not surprising that not all teachers effectively micro-adapt (Clark & Yinger, 1977).

**Research versus practice.** It appears that microadaptation has been largely neglected by research (Corno, 2008; Nuthall, 2004). On the one hand, this is because micro-adaptation is by definition unplanned and is therefore hard to track (Corno, 2008; Corno & Snow, 1986). On the other hand, this is due to the different ways in which research is viewed. Researchers seem to view their findings as information which should directly shape teaching practice, while teachers perceive research as something which must be heavily tailored to meet their own needs (Hiebert, Gallimore, & Stigler, 2002; Kennedy, 1999; Levin & O'Donnell, 1999) or something that takes away from valuable teaching time (Corno, 2008).

The fact that instructors perceive research in such a fashion may be why instructors are often reluctant to change their teaching approaches based on research recommendations (Randi & Corno, 1997). Therefore, some researchers (Corno, 2008; Nuthall, 2004) have suggested that research should (1) examine what kinds of tailoring actually take place in classrooms, (2) experimentally assess any resulting impact, and (3) document the specific linkages between performance changes and activities.

Corno (2008) and Nuthall (2004) note that the link between research and practice is a bi-directional need, so to speak. Research should be informed by the constraints of classroom practice, taking into account time and resource constraints as well as the appropriateness of certain methods for certain domains. Conversely, practice should be well-informed by empirical assessments of method efficiency/effectiveness. It is often difficult if not impossible for teachers to informally and accurately assess the progress of all students; often they must rely on a subsample, given time and memory constraints (Alton-Lee, Nuthall, & Patrick, 1993). Furthermore, teachers are often looking for obvious indicators which might impede instruction, such as attentiveness or boredom (Corno, 2008), which is not the same as observing actual, direct indicators of performance.

**Research implications.** In sum, it appears that there is an important gap in the academic literature. On the one hand, the central role of microadaptation in tailoring is seen as self-evident. On the other hand, there has been little systematic or principled integration between pedagogical practice and empirical assessments of pedagogical methods. This gap also exists in U.S. Army contexts.

## **Learning Styles**

The educational literature is rife with references to learning styles. As Curry (1990) noted, some learning style theorists seem to think of learning styles as akin to learning preferences: a stable predilection for information presented in a specific manner (e.g., pictorially,

verbally, or aurally). For tailored training purposes, however, the most salient conceptualization of learning styles revolves around the “meshing hypothesis” (Pashler, McDaniel, Doug, & Bjork, 2009). The meshing hypothesis claims that matching learning conditions to learning styles enhances performance, and mismatching learning conditions and styles suppresses performance.

For example, say that a group of students have completed a learning style inventory. Half are classified as ‘visual’ learners and half as ‘auditory’ learners. Half of the visual learners are assigned to a condition in which information is presented visually or aurally. The same procedure is applied to the auditory learners. A common test is then administered to all of the students to measure comprehension of the material. The meshing hypothesis predicts two things. First, that aural learners in the aural condition will outperform aural learners in the visual condition. Second, that visual learners in the visual condition will outperform visual learners in the aural condition.

It is our considered view that investing in learning style measures will not yield appreciable tailored training results. First, there are long-standing criticisms of learning style measures regarding inadequate validation (Curry, 1990) including failure to assess competing factorial models, tendency to re-label familiar constructs, and failure to establish sufficient reliability (Coffield, Moseley, Hall, & Ecclestone, 2004).

However, the most important reason for our skepticism towards learning style research is that solid evidence for the meshing hypothesis is lacking. Pashler et al. (2009) state that there are few studies which even attempt to demonstrate that performance interacts with learning preferences (i.e., that the meshing hypothesis is true). Of those that do attempt to do so, some results falsify the meshing hypothesis. In short, “...at present, these negative results, in conjunction with the virtual absence of positive findings, lead us to conclude that any application of learning styles in classrooms is unwarranted” (Pashler et al., 2009, p. 112).

### **Aptitude-Treatment Interactions**

Beginning in the 1960s, Cronbach and Snow (Cronbach & Snow, 1977; Snow, 1992) began to examine aptitude-treatment interactions (ATI). An ATI is present when the relationship between an aptitude (an individual difference variable) and performance varies from condition to condition (Cronbach & Snow, 1977). ATI suggest another approach to tailoring training: placing individuals or groups in specific conditions based upon level of aptitude. The ATI concept brings to the fore one aspect of our definition of tailored training – it addresses critical individual differences in learners.

For approaches to tailoring based on ATI findings to be effective, at least two conditions must be satisfied. First, there must be evidence demonstrating a significant relationship between one or more aptitudes and performance. Second, there must be evidence of an interaction between one or more aptitudes and the training condition (i.e., there must be evidence of the sort that Pashler et al. [2009] did *not* find for learning styles).

With regards to the first condition, we note that decades of individual differences research (Corno, 1992; Jensen, 1998) yield the conclusion that the most powerful and

generalizable predictor of performance is general mental ability. This remains the case despite the fact that "...innumerable studies have tried to raise correlations by weighting measures of personality or motivation alongside ability in the prediction, and their hopes of improving prediction were generally disappointed. Positive findings were scattered and inconsistent" (Corno, et al., 2002, p. 105)

ATI research declined in the late 1970s because of these inconsistencies. Pelligrino, Baxter and Glaser (1999) attributed the inconsistencies in ATI findings to the mismatch between the established tests of aptitude derived from differential psychology approaches to personnel selection and classification, and the learning and performance settings investigated in experimental psychology (for psychology related discussion, see Cronbach, 1957). On the other hand, Shute (1992, 1993) attributed the decline and inconsistencies in results to the "noisiness" in the data that resulted from extraneous uncontrolled variables which made interactions hard to find and to interpret.

With regards to the second condition, Snow (1992) states that "Measures of general ability...reflect an important aspect of aptitude and show many ATI but interact especially when one treatment can be characterized as highly structured, complete, and direct and another can be characterized as relatively unstructured, incomplete, and indirect" (pp. 11). These conclusions were also stated by others (e.g., Jonassen & Grabowski, 1993; Pellegrino, et al., 1999).

In the late 1970s and beyond, the study of individual differences expanded to include comparison of experts and novices in diverse fields, changes in standardized test procedures to assess reasoning and complex understanding, and the development of information processing theories. "The study of expertise led to an alternative approach to the study of individual differences ... focused on attained knowledge and related cognitive processes that are the object of deliberate instruction, practice and learning" (Pellegrino, et al, 1999, p. 317). The authors concluded that current ATI research, termed "second-generation ATI," is based on contemporary information-processing theories. Further, in second-generation ATI research individual differences measures are often expanded to include general prior knowledge and/or ability to learn new tasks that tap processes assumed to underlie the processes required by the treatment condition itself.

Additionally, general mental ability primarily affects performance indirectly through the gaining of prior knowledge (Pashler et al., 2009; Schmidt & Hunter, 1992). Therefore, the general conclusion that we derived from the ATI research is that the most promising approach for tailoring training in military settings is to assess prior knowledge, and use high-structured techniques for low knowledge individuals and low-structured techniques for high prior knowledge individuals. Most of this section deals how we arrived at this conclusion. First, however, we examine how ATI have been analyzed and the general forms such interactions can take.

**Analysis of ATI.** There are two major approaches to analyzing ATI. First, one can measure aptitudes, analyze that information, and use that information to assign individuals to different conditions. Second, one can measure aptitudes, assign the group as a whole to a specific condition, and then analyze post-hoc the aptitude information and assess its relationship

to performance. As an example of the first approach, a researcher can measure the prior knowledge of math students and use a median split of the pre-treatment scores to assign individuals to “high” or “low” prior knowledge conditions. Equal numbers of participants from each knowledge condition can then be randomly assigned to one of two treatment conditions, resulting in a 2 (prior knowledge) by 2 (treatment condition) experiment (see Kalyuga & Sweller, 2004). If a significant analysis of variance (ANOVA) interaction term is present, then a significant ATI is present.

The second approach is typically used when either (a) the group is large enough that it can be expected to display enough range in the aptitudes of interest (Goska & Ackerman, 1996) or (b) the research is dealing with intact groups, as in school research (Cronbach & Snow, 1977). The presence of an interaction is then assessed by using regression techniques (i.e., testing for differences in correlations, interaction terms, or slopes).

In our review of the literature, we found that the former approach is typical of recent articles, and the latter is typical of older articles. We will make clear which approach was used when discussing specific findings.

**Types of ATI.** Possible individual difference by treatment interactions—like any other kind of interaction—can vary in shape. Notional representations (not tied to any particular finding or data set) are displayed in Figure 1 (see Cronbach & Snow, 1977; Smith & Sechrest, 1991). We realize that this selection is not exhaustive. However, we chose these interactions because the training recommendations drawn from each are quite different.

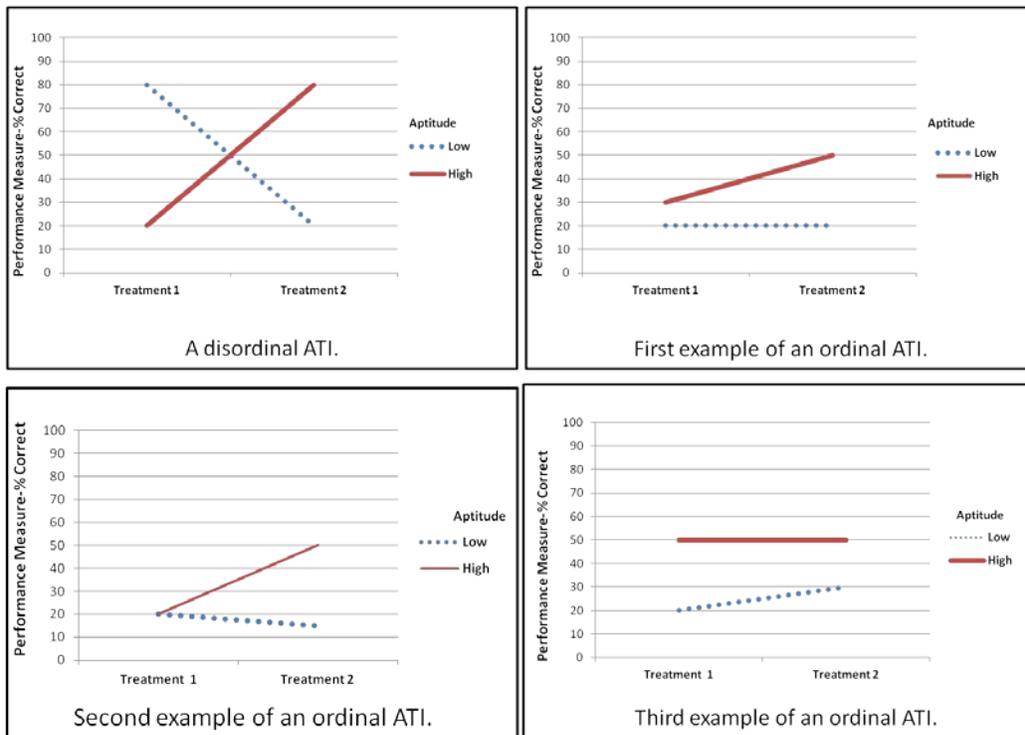


Figure 1. Notional Aptitude-Treatment Interaction.

The top left graph in Figure 1 is a disordinal interaction in which the lines cross. This simply means that there is a reversal in the rank of means as treatment conditions change. In Treatment 1, low aptitude individuals outperformed high aptitude individuals. In Treatment 2, the pattern was reversed. When such an interaction involves the aptitude of prior knowledge, the interaction is termed an expertise reversal effect (ERE). The core concept of the expertise reversal effect literature is that treatments which improve performance for novices can become detrimental as expertise (that is, prior knowledge) is gained. Conversely, treatments which are detrimental for novices become beneficial as expertise is gained.

The remaining interactions are ordinal. In ordinal interactions, the rank order of means remains the same across treatment conditions, but the delta between the means changes, either decreasing or increasing. The pattern of these changes varies from interaction to interaction. In the top right figure, the interaction derives from the fact that low aptitude individuals performed the same across the two treatment conditions, and in both cases performed more poorly than high aptitude individuals. In Treatment 2, however, the scores of the high aptitude individuals increased relative to their performance in Treatment 1. In the bottom left figure, Treatment 2 helps the performance of high aptitude individuals and hurts the performance of low aptitude individuals. In the bottom right figure, the performance of the high aptitude individuals remains the same across the treatments, but low aptitude individuals perform better under Treatment 2 than under Treatment 1.

The training recommendations drawn from each interaction differ. In the case of the top left figure, the recommendation is clear: if possible, assign high aptitude participants to Treatment 2 and low aptitude participants to Treatment 1. If only one treatment can be used for all trainees, then the decision maker will be forced to make a choice of which group of trainees (low or high aptitude) will take the hit. In the case of the top right graph, the recommendation would be to use Treatment 2. Treatment 2 does not harm the performance of low aptitude participants relative to Treatment 1, and significantly improves the performance of high aptitude individuals. In the case of the bottom left figure, the recommendations cannot be made in a vacuum. If the goal is to ensure some minimal performance standard, then Treatment 1 might suffice for aptitude groups. If the goal is to maximize performance, then low aptitude individuals should be assigned to Treatment 1 and high aptitude individuals to Treatment 2. However, if time and resources do not permit the use of two different conditions, it might be prudent to use the Treatment 1 condition for all participants as the performance of low aptitude individuals would not be lowered. In the bottom right graph, the training recommendation is again clear: use Treatment 2.

We wish to stress that the inclusion of these examples is merely a way to illustrate what kinds of recommendations might be drawn from different significant interactions. We make no judgment on which interactions are the most common. Both ordinal and disordinal interactions can have practical implications.

**Roadmap.** The roadmap below (Figure 2) gives the reader an idea of how we will proceed in our ATI discussion. The ATI discussion will consist of four subsections, corresponding to ‘aptitudes’ (the first column of the graphic), ‘treatments’ (the second column), ‘interactions’ (the third column), and ‘explanations’ (the fourth column). Boxes and connections

with dashed lines are used to indicate that the concept or its relationship to ATI research is not firmly established.

In the aptitudes section, we briefly recap why certain aptitudes (with dashed lines) are not seen as fruitful bases for tailoring training. We present evidence supporting the contention that general mental ability (GMA) and experience impact performance strongly but indirectly through the development of prior knowledge. The result is that prior knowledge becomes the most direct and therefore the most powerful predictor of performance.

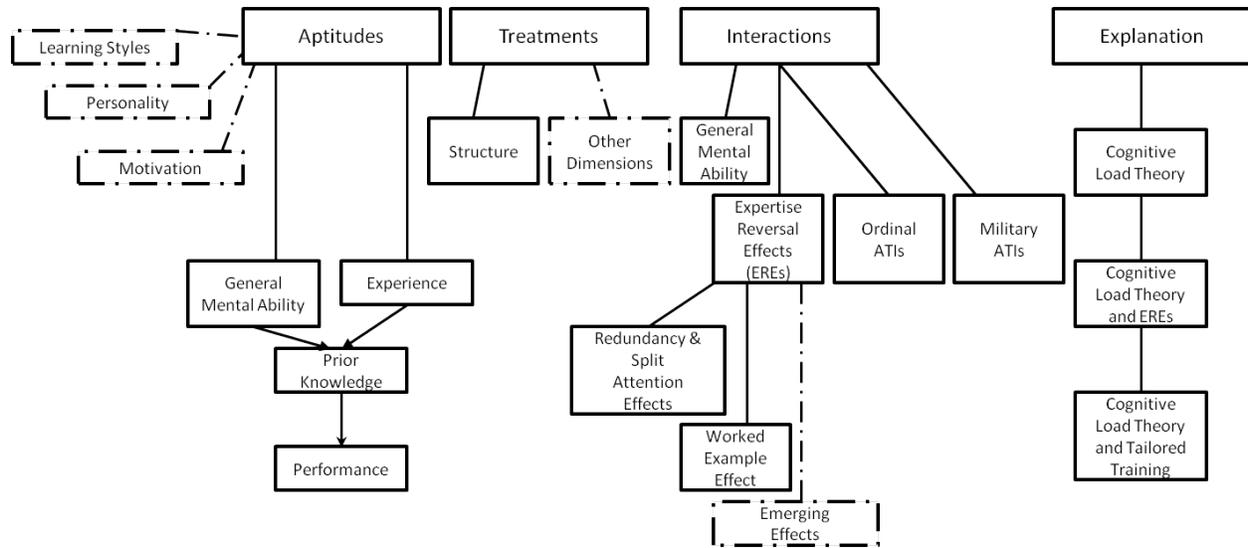


Figure 2. Roadmap of the aptitudes, treatments, interactions, and explanation sections.

In the treatments section, we discuss how the concept of ‘structure’ is at the core of the majority of treatment manipulations in ATI. We also briefly discuss some under-researched dimensions (hence the dashed lines) along which treatments may be varied.

In the interactions section, we discuss an example of GMA interacting with transfer tasks and then spend the bulk of the section detailing EREs. EREs essentially seek to understand how differing levels of prior knowledge interact with treatments. We then discuss three of the more replicated ERE findings (redundancy, split attention, and worked example effects). We also briefly summarize three interesting but not well-researched findings (hence dashed lines) under ‘emerging effects’. We discuss some examples of ordinal ATI to underscore the point that despite the research literature emphasis on EREs, it is by no means given that an ATI will be disordinal. We round out the interaction section by providing some examples of ATI found with military subject matter and populations.

In the explanations section, we summarize the nature of the interactions just examined. We also present a theoretical explanation for the role of prior knowledge in ATI known as cognitive load theory. We end by discussing some of the resulting implications for tailored training.

**Aptitudes.** A variety of aptitude-performance relationships have been examined. As indicated earlier, the consensus is that basing prediction of performance upon ability measures alone does about as well as using measures of ability plus learning style, personality, and motivation. Therefore, given the current state of knowledge, it seems that there are really only three aptitudes which warrant further investigation: general mental ability (Snow, 1992), experience (Schmidt & Hunter, 1992), and prior knowledge (Snow, 1992). As will be seen, these three aptitudes can often be reduced to just one: prior knowledge. We begin with a discussion of the evidence for general mental ability as a performance predictor. But first, we acknowledge that Shute's (1992; 1993) approach to assessing individual differences in terms of the information processing requirements of the tasks to be learned (e.g., basic associative learning ability and/or working memory tasks plus knowledge) is another approach that could be considered and may have promise.

***General mental ability as a performance predictor.*** A reasonable definition of GMA is provided by Gottfredson (1998): GMA is that dimension tapped to a greater or lesser extent by all intelligence tests, regardless of content (Gottfredson, 1998; Jensen, 1998; Ree, Carretta, & Teachout, 1995; Schmidt & Hunter, 1992; Teachout, 1995; Thorndike, 1985). We chose this definition because it makes clear an important point: while a test may have been designed to measure a different construct (e.g., fluid intelligence, crystallized intelligence, or verbal intelligence), a common finding is that the test often overlaps with other ability measures primarily because it also taps GMA (Carroll, 1993). Because the content of GMA measures and domain knowledge measures may not overlap at all, GMA is conceptually distinct from measures of prior domain knowledge.

This pattern of results has led researchers to state that the strong association between GMA and "performance in a wide variety of domains is one of the most consistent findings in our field" (Gully & Chen, 2010, p. 9). This holds true in both civilian and military contexts (Schmidt & Hunter, 1992, 1993). There is thus ample evidence that GMA predicts performance.

***Experience as a performance predictor: direct or indirect?*** Schmidt, Hunter, and Outerbridge (1986) noted that experience is an obvious predictor of job performance, although the relationship between experience and job performance can be direct or indirect. In the direct case, the relationship between experience (opportunity to learn) and job performance may be unaffected by moderating variables. In the indirect case, the impact that experience has upon job performance may be contingent upon other variables. Two possible moderating variables are GMA and prior knowledge, which can be defined as information, facts, and procedures required for successful performance (Chen & Paul, 2003; Palumbo, Miller, Shalin, & Steele-Johnson, 2005; Schmidt et al., 1986).

To assess whether experience impacts job performance directly or indirectly, Schmidt, Hunter, and Outerbridge (1986) applied path analysis to four military studies previously reported by Vineberg and Taylor (1972). Those military studies assessed Soldiers ( $N = 1,474$ ) from four U.S. Army jobs (armor crewman, armor repairman, supply specialist, and cook) in terms of five variables. The salient variables reported across the four studies were job experience, general mental ability, job knowledge, and work sample performance. The variables were operationally defined as follows. Job experience was simply the number of months on the job. The test of

general mental ability was the Armed Forces Qualification Test (AFQT). The measures of job knowledge were simply referred to as paper-and-pencil tests of job-relevant facts and procedures. Detailed descriptions of the work sample tests and supervisory rating instruments were not provided.

The studies met specific criteria set by the authors. First, the Soldiers had to be neither too experienced nor too inexperienced, as either situation could artificially mask the impact of ability upon performance (whether measured by work samples or supervisory ratings). Second, the correlation between ability and experience had to be low, or causal indeterminacy would result. Third, the jobs were of moderate complexity, as too little or too great complexity can complicate causal interpretation of correlation matrices.

Path analyses of the results indicated that the effects of general mental ability and job experience upon work sample performance were largely indirect in nature. Ability and job experience directly impacted job knowledge, which in turn impacted work sample performance. That is, the direct correlation between ability and work sample performance was close to zero. However, ability was significantly correlated with job knowledge which in turn was correlated with work sample performance. Similarly, job experience was only weakly related to work sample performance but significantly related to job knowledge, which in turn was significantly related to work sample performance. Thus, job knowledge was a much better predictor of work sample performance than were ability and job experience. (For a similar pattern among prior knowledge, experience, and performance in a military course context, see Schaefer, Blankenbeckler, & Lipinski, 2011.)

Similar findings have been found with different jobs and participants. Borman, White, Pulakos, and Oppler (1991) examined first-tour Soldiers ( $N = 4,362$ ) from nine U.S. Army jobs. The nine jobs were cannon crewman, vehicle mechanic, administrative assistant, infantryman, tank crewmember, radio operator, motor transport operator, medical care specialist, and military police. The sample was relatively homogeneous in terms of experience, with most participants having 36-40 months in the Army. Once again, general mental ability and hands-on task proficiency were measured.

The results parallel those of Schmidt et al. (1986). The direct effect of ability upon task proficiency was small ( $r = .13$ ) compared to the indirect correlation, derived by multiplying the correlation between ability and job knowledge ( $r = .66$ ) and the correlation between job knowledge and task proficiency ( $r = .66$ ; indirect  $r = .32$ ). Again, the impact of ability upon job performance seems to be mediated by job knowledge, and job knowledge is a much better predictor of performance than ability.

Borman, White, and Dorsey (1995) used the same measures as the Borman, White, Pulakos, and Oppler (1991) study, while including a few more variables in an attempt to increase the variance explained by the model. The results exhibit the same pattern as shown before: the direct impact of ability upon task proficiency was comparatively weak while the impact of ability upon job knowledge was comparatively strong. In turn, job knowledge significantly predicted technical proficiency. Once again, job knowledge was the most direct and hence strongest predictor of task proficiency.

Ree, Carretta, and Teachout (1995) attempted to replicate the above pattern with training performance as the variable. The authors measured general mental ability and prior job knowledge and examined the impact of these variables on acquiring subsequent job knowledge during training as well as work-sample performance during training. Participants ( $N = 3,428$ ) United States Air Force officers completing a 53-week pilot training course between 1981 and 1993. Participants had already been screened for officer commissioning based in part on their Air Force Officer Qualifying Test. Therefore, the measures of general mental ability and job knowledge were evaluated by gathering scores from the appropriate subtests. Measures of job knowledge acquired during training were derived from classroom grades. Work samples were composed of blocks of flying time completed by participants. After each flight, performance was evaluated via work-sample tests called check flights.

As there were several measures of job knowledge acquisition during training, we provide here only an overall narrative summary of the data pattern. The direct paths between general mental ability and work-sample performance were near zero in all cases. General mental ability exercised its influence upon work-sample performance almost entirely through its influence upon prior job knowledge. The role of prior knowledge, however, must be examined more closely. In this study, there were ongoing assessments throughout training of both job knowledge and work performance (piloting skills). As earlier blocks of training are foundations for later blocks of training, correlations between measures of prior knowledge and job performance would be expected to be larger if the measures are concurrently administered. This is in fact what happened. Prior knowledge did a better job predicting earlier measures of job knowledge than later ones, and prior knowledge did a relatively poor job of predicting later work sample performance. In contrast, the best predictor of the second flight check test was performance on the first flight check test ( $r = .92$ ).

This finding suggests something of importance. Prior knowledge can often predict performance in a narrow domain better than general mental ability. However, skills are not static. When additional training takes place, military task performance appears to follow the so-called 'power law' of practice, wherein the largest improvements in performance take place early in training but gradual, continued improvement takes place as training proceeds (Dyer, 2004). In such situations, a better predictor might be a variation of the task itself (Hailikari, Neygi, & Komulainen, 2008; Palumbo, Miller, Shalin, & Steele-Johnson, 2005; Regian & Schneider, 1990;).

***Prior knowledge as a performance predictor.*** To sum up, prior knowledge is often a better indicator of job performance than general mental ability. The theoretical explanation for this is that prior knowledge captures differences in both job experience and general mental ability (Schmidt & Hunter, 1992, 1993). Prior knowledge is in effect the combination of capacity to learn (general mental ability) wed to the opportunity to learn (experience). Consistent with this interpretation is the observation that as one variable becomes more homogeneous, the explanatory value of the other variable is enhanced. Schmidt and Hunter (1992) note that when the effects of experience are statistically controlled for, the correlation between general mental ability and performance increases. When differences in general mental ability are controlled, the correlation between experience and performance increases. Prior

knowledge has also been found to be a major predictor in academic performance as well (Dochy, Segers, & Buehl, 1999; Jonassen & Grabowski, 1993; Shapiro, 2004; Tobias, 1989). Carroll (1967) put this relationship and its implications for instruction in slightly different terms advocating that you should teach at the appropriate level on one or more learning curves, based on a thorough and accurate diagnosis of the student's knowledge status. Glaser (1984) also stressed the important of domain-specific prior knowledge as a major factor impacting an individual's ability to acquire additional knowledge, and to think and solve problems.

Thus, we seem to be on firm footing when we recommend the use of prior knowledge as a predictor of early training performance. As noted above, however, if training and attendant performance improvement takes place over a long period of time, ongoing assessments of performance may be necessary. If, for example, an incoming Soldier was administered a prior knowledge test at the beginning of a course six weeks earlier, better prediction of performance on an imminent criterion might be achieved by assessing the Soldier's current knowledge status.

We have thus satisfied the first precondition of ATI-driven tailored training: we have isolated several aptitudes (GMA, experience, and prior knowledge) which significantly predict performance. In many situations, this set can be further reduced to just one aptitude: prior knowledge. (This is assuming that prior knowledge is in fact correct. For a discussion of how to minimize the negative impact of inaccurate prior knowledge, see Dochy et al., 1999). Now we can turn to a consideration of the types of treatments which ATI research has examined.

**Treatments.** This section is by far the shortest of the four displayed in Figure 2, and for a very good reason. Namely, that "The need for a better conceptualization of treatments...is one of the most persistent issues in ATI research" (Jonassen & Grabowski, 1993, p. 28). In other words, much attention has been paid to the conceptualization of aptitudes, but not nearly as much has been paid to the treatment manipulations. It could be that if a review focusing on just the ATI research involving prior knowledge were conducted, a consistent pattern would be revealed. However, we do not know of any such existing review.

By far, the most common treatment manipulation in ATI research involves "structure." We first examine the concept of structure, and then examine other treatments.

**Structure.** As noted above, Snow (1992) stated that the most common ATI involves GMA and treatments which vary in structure, directness, and completeness. Others have described structure slightly differently. Berliner and Cahen (1973) used the terms of teacher-centered, didactic, conforming and lecture to describe structure, and the terms of student-centered, flexible, independent study, and discussion method to describe unstructured. Note that these terms seem to apply to classroom settings as opposed to experimental settings. Jonassen and Grabowski (1993) discussed the structure dimension in terms of degree of instructional support compared to placing the burden of information processing on the student. Pellegrino et al. (1999) referred to treatments that put the burden of organization (such as discovery learning) on the student as low structure, compared to treatments which provide elaborated, complete materials and direct instruction as high structure. While all these terms are suggestive, they do not clearly convey the nature of structure treatments. Therefore, it behooves us to consider some examples.

Goska and Ackerman (1996) used a criterion task composed of cognitive and procedural demands. One group of participants was trained on both demands, while the other participant group was trained only on the procedural tasks. All participants were then tested on the criterion task, and thus had to be able to meet both cognitive and procedural demands. Arguably, this treatment manipulation maps well onto the (un)structured, (in)complete, and (in)direct dimensions mentioned by Snow. First, the training obviously differed in structure in that whatever learning of the cognitive demands took place in the second participant group, it was not systematically presented and varied. Second, the training obviously differed in completeness in that the second participant group received training on just the procedural demands. Third, the authors argued that any learning of the cognitive demands which took place was largely indirect, as successful performance in training was not contingent on learning the cognitive demands.

Shapiro (2004) compared detailed text material with sparse text material. The sparse text simply described the sequence of historical events (low structure), while the detailed version had more instructional support as it contained information on the reasons for the events (high structure). Ross and Rakow (1981) compared computer based training which controlled the number of practice problems in mathematics based on pre-test scores (high structure/instructional support) to a learner controlled mode where the learner determined the number of practice problems (low structure/instructional support). Using college calculus students, Pascarella (1978) compared lecture (low instructional support) to the PSI (personalized system of instruction – high support) which allowed self-pacing, offered tutorial sessions, had detailed self-study guides and optional problem solving sessions.

Dyer, Singh and Clark (2005) compared five different computer-based modes to train Soldiers on map-related digital skills. Drop-down menus were used to initiate the map functions. The two highest structured conditions involved solving practice exercises, but one condition was a traditional lesson followed by practice exercises for each topic and the other just required Soldiers to solve exercises (no formal lesson). None of the other three conditions required Soldiers to solve exercises and had less instructional support. One condition only had the lessons but the Soldier could explore the map after the lesson. The other two conditions were learner-controlled but varied considerably. The self-select condition let a Soldier choose the mode(s) of instruction for each topic- take a lesson, solve exercises and/or explore the map. The least structured condition simply let the Soldier explore the map using the available menu selections.

***Other dimensions.*** The majority of treatment manipulations used in ATI research correspond to the high vs. low structure conditions discussed by Snow (1992). This may be partly due to the fact that certain treatment manipulations are associated with relatively weak predictors of performance, such as learning style measures. In many cases, the robustness of any interactions between the proposed dimension (e.g., cognitive styles) and treatments is hard to assess as there appear to be few if any existing meta-analyses of such interactions.

Berliner and Cahen (1973) identified three major categories of treatments, in addition to structured and unstructured, which had been examined: inductive vs. deductive, questioning strategies (adjunct questions), and subject matter (e.g., phonics vs. whole word approach in reading; old vs. new math). The impact of verbal versus spatial materials has been investigated extensively (Snow, 1976), but, the expected interactions have not been found. However, it is

noted that there are some specific military contexts where the visual and verbal media dimensions are relevant (see Dyer, Gaillard, McClure & Osborne [1995] research on an unaided night vision training program). In addition to treatments that varied structure, Snow (1976) cited variations in the sequence of instructional materials, the pace of training, and the use of demonstrations or models. In a review of technology-based adaptive instructional procedures, Durlach and Ray (2011) found that variations in feedback were common (e.g., types of hints when errors were made, immediate vs. delayed, expository feedback vs. question type feedback, and automated feedback after errors and uncertain responses vice only after errors). They also found treatments that varied the fading of worked examples, changed the type of drill and practice (systematic vs. random), and adapted the difficulty level of the content to the performance of the student. Shute (1993) varied the number of practice items.

Another treatment dimension is time on task. One way of tailoring instruction is to accommodate to differences in the rate at which individuals learn, which is commonly done in instructional technology applications. Here individuals typically progress at their own rate, but the basic content of such programs may not vary greatly beyond the type of feedback provided (e.g., Gibbons & Fairweather, 2000; Kulik, 2003; Kulik & Kulik, 1991).

**Interactions.** As noted earlier, there are two conditions which must be met before ATI tailored training can be successful. First, the proposed aptitude(s) must reliably and significantly predict criterion performance. Secondly, there must be demonstrable interactions between the aptitude(s) of interest and treatment conditions. Given the discussion in the preceding aptitude and treatment sections, it will not surprise the reader to find that we focus our discussion on interactions involving general mental ability (GMA), prior knowledge, and structure.

***Interactions involving GMA.*** There were hints early on that GMA interacted with treatment conditions. For example, in Jones (1948) students completed an intelligence quotient (IQ) test (a proxy variable for general mental ability—see Jensen, 1998) and were then assigned to either a control group or an experimental group. In the control group, only minor changes were made to instructional methods and no changes to instructional materials. In the experimental group, teachers were encouraged to tailor instructional methods and materials to meet the achievement level, needs, interest, and rate of progress to their students. All students were measured in a pre- and post-intervention fashion in three areas (spelling, reading, and arithmetic) to assess gain scores. In all three areas, the gain scores were larger in the experimental group. This is, of course, to be expected. However, the most germane finding for our purposes is that the gain scores in the experimental group varied with estimated IQ level. That is, significant gain scores were obtained only for students with IQs below 110. (Thus, this appears to correspond to the bottom right graph in Figure 1.) Once again, the association between performance and general mental ability varied—in this case, as both a function of treatment and level of general mental ability.

Similarly, Cronbach (1957) discussed an experiment in which students learned material from either text or film. Performance on a subsequent test over the material was then correlated with a measure of GMA. The correlations varied between the two conditions from .30 (text) to .77 (film).

Goska and Ackerman (1996, Experiment 2) provide another example of GMA by treatment interaction. In that experiment, two groups of individuals completed measures of GMA. One group was trained on both the cognitive and procedural demands of a flight simulator task (near transfer), while the other group was trained on just the procedural demands (far transfer). The expectation was that GMA would be a better predictor of performance for the second group than the first. As the authors put it “Students had the opportunity to learn about the rules for the task, but that was not required by the training task. Because the learning opportunity was provided but not required, we expected that it would result in a greater transfer advantage for higher ability students, whose attention is not totally consumed by the requirements of the training task” (Goska & Ackerman, 1996, p. 254). This hypothesis was partially supported. One of the performance measures (successful plane landings) showed the expected correlational pattern with GMA. Namely, the correlations between GMA and successful landing were consistently and significantly larger for the group that received training on the procedural demands only, compared to the other group.

As Goska and Ackerman note, the training recommendations from this are somewhat equivocal. One interpretation is that higher aptitude individuals are somehow able to extract more ‘context free’ skills than lower aptitude individuals, and that more training might not help. However, another interpretation would be that higher aptitude individuals simply learn faster, and that providing the lower aptitude individuals with more training could close the gap.

***Expertise reversal effects.*** Given the fact that general mental ability largely impacts performance indirectly through the acquisition of prior knowledge, it is unsurprising that many of the ATI found in the literature involve prior knowledge—primarily EREs. An ERE is present when a treatment which is beneficial to novices becomes deleterious as expertise is gained. Conversely, treatments which were originally detrimental to novices become helpful as experience is gained (Kalyuga, Ayres, Chandler, & Sweller, 2003). An ERE is thus a disordinal interaction. However, this does not mean that all ATI involving prior knowledge (or ATI in general) are disordinal, just EREs. Examples of (largely) ordinal interactions follow the discussion of EREs.

Differences in prior knowledge can either be measured (i.e., they are pre-existing differences) or induced (e.g., through number of learning trials). Prior knowledge can be measured by examining persons who can be expected to vary in experience in systematic ways. For example, assume that the domain of interest is mathematics. Persons who have had more mathematics classes can be assumed to be higher in prior knowledge. Alternatively, one can administer a prior knowledge measure (either one specifically targeted to the content or utilizing a domain-relevant standardized achievement test). The typical procedure is then to impose a high/low prior knowledge split upon the test scores—based either upon the median or the average score—followed by assignment of participants to treatment conditions.

Prior knowledge can be induced when novel laboratory tasks are used. Because the task is assumed to be new to all participants, expertise can be operationally defined as the number of trials with that task. Thus, this approach relies upon repeated-measures designs. The goal would then be to show that treatments which are beneficial in earlier blocks of trials become less helpful and then (given enough trials) detrimental as experience accumulates. Conversely,

treatments which are detrimental in earlier blocks of trials become more helpful as experience is gained.

Because there are, to our knowledge, no meta-analyses of ERE findings, we first provide a narrative summary of some ERE studies, and then conclude the ERE section with a table which enables the reader to get a ‘birds-eye’ view of this domain. The narrative summary focuses most on those studies whose findings appear to be the most robust, and briefly outlines some ERE findings which are interesting but need to be replicated. The table conveys the types of domains in which EREs have been found and how prior knowledge was assessed.

*Redundancy and split attention effects.* These effects are described in the same section because they are complementary in nature: each refers to a specific portion of a full ERE (i.e., a disordinal interaction). The core idea behind the redundancy effect is as follows: commingling new information with information that is ‘old hat’ to an individual retards that individual’s performance. The ‘old hat’ information is redundant for that person (Mayer, Heiser, & Lonn, 2001). Therefore, some mental resources are needlessly engaged in processing and/or filtering out the redundant information. Such an individual would perform better if the new and the old information could be parceled out (i.e., put on separate pages or displayed on different computer monitors). Thus, the individual could just focus on the new information, avoiding the redundancy effect.

The core idea behind the split-attention effect is just the opposite. Say that the same set of materials mentioned in the prior paragraph is administered to individuals for whom all of the information is new. Then the opposite pattern of results should be seen. Presenting all of the information together would be more efficient. For this set of individuals, the information is not redundant. Therefore, displaying the information on separate pages or monitors would require this set of individuals to split their attention (hence, split-attention) effect across pages or displays.

In short, prior knowledge interacts with how the information is displayed. For high prior knowledge individuals, information should be parceled out into redundant and non-redundant (separate pages or displays). If such information is combined, performance will suffer. For low prior knowledge individuals, information should be combined. If low prior knowledge persons are required to mentally integrate information across separate pages or displays, performance will suffer (Florax & Ploetzner, 2010).

For example, consider Yeung (1999), who was interested in how EREs might impact the text comprehension and vocabulary acquisition of Hong Kong Students of English as a Second Language (ESL). However, these findings also underscore the need for understanding the subtleties of task demands. First, consider what text comprehension requires. A precondition of text comprehension is vocabulary acquisition. One cannot understand a body of text unless one understands the words used in the text. So persons with a poor grasp of English vocabulary (‘low prior knowledge’) might perform better in a single display condition consisting of text and embedded vocabulary definitions than if those information sets were displayed separately. Conversely, persons with a good grasp of English vocabulary (‘high prior knowledge’) would be expected to display the opposite pattern, performing better in the separated display condition

because the vocabulary definitions would be superfluous for them. These expectations were borne out.

Now consider the same basic scenario but switch the task from text comprehension to vocabulary acquisition. In this case, individuals low in vocabulary knowledge might perform better in the separate display condition as the load imposed by the text is extraneous to the narrower task of vocabulary acquisition. However, individuals high in prior vocabulary knowledge might perform better in the single display condition as they might be able to ‘glean’ clues from the surrounding text as to the nature of the word. Again, these expectations were borne out.

Broadly similar results were also found in Yeung, Jin, and Sweller (1998), and in a repeated measures design with mechanical apprentices (Kalyuga, Chandler, & Sweller, 1998). Redundancy effects have also been found with purely textual materials (McNamara, Kintsch, Singer, & Kintsch, 1996, and Sweller, van Merriënboer, & Paas, 1998). Redundancy effects and split-attention effects seem to be relatively robust expertise reversal phenomena. The take away message for tailored training is that individuals who vary widely in prior knowledge require different materials to improve performance efficiently and effectively. For example, providing users low in prior knowledge information on a Mission Command System (formerly known as Army Battle Command System) might require an integrated display linking buttons to descriptions of associated functions. For users high in prior ABCS knowledge, such information might better be provided in separate panels with linkages made only if explicitly requested by the user.

*Worked example effect.* One way of describing this effect is to contrast worked examples with problem solving. A worked example is an example problem that makes explicit the series of steps involved in solving a problem. The burden is placed on the instructor or instruction delivery system (e.g., computer) rather than the learner. At the other extreme is problem solving. Here, problem solving is operationally defined as presenting the learner with the problem alone. The basic idea behind the worked example effect is that worked examples help low prior knowledge individuals but hinder high prior knowledge individuals. Conversely, high prior knowledge individuals do better with problem solving than low prior knowledge individuals. Research has indicated that worked examples benefit low prior knowledge individuals but hinder high prior knowledge individuals (Tuovinen & Sweller, 1999). In a within-subjects design, Kalyuga, Chandler, Tuovinen, and Sweller (2001) presented trade apprentices with familiarization training on writing programs for relay circuits and then presented either worked examples or problem solving. The same comparison was made again after two more training sessions, and one final time after yet more training. With some qualifications, the predicted effect was found when considering the first and last manipulations. Similar results were found by Kalyuga, Chandler, and Sweller (2001). The worked example effect is perhaps the clearest example of variation in structure.

*Emerging effects.* In this section, we discuss three different effects which fall under the ERE umbrella but which have not been replicated enough for us to repose much confidence in the results. The first of these might be dubbed the technology effect. Clarke, Ayres, and Sweller (2005) posited that familiarity with technology used to present information can itself be a

variable of interest. Clarke et al., (2005) used spreadsheets to present mathematical concepts to fourth-grade students. Based on school records, the students were roughly equivalent in math knowledge. The authors posited that prior knowledge of spreadsheets would itself interact with treatment. Students self-rated themselves on spreadsheet familiarity, and then were either to a 'sequential' or 'simultaneous training condition. In the sequential training condition, students were first trained on spreadsheet use and then were exposed to math concepts via spreadsheets. In the simultaneous training condition, the spreadsheet use training was omitted. Results indicated that, as expected, students with little experience with spreadsheets (low prior knowledge) performed better under the sequential training rather than the simultaneous training.

The second of these effects might be dubbed the imagination effect. Leahy and Sweller (2005) assigned fifth-graders to an 'imagination' versus 'practice' condition (there was another manipulation involved which is not of central interest, and is thus omitted). The students used temperature graphs to solve problems. In the 'imagination' condition, students were given examples of imagination by the experimenter and were then given access to the instructions alone while they imagined solving problems using the temperature graphs. In the 'practice' condition, participants were given simultaneous access to the instructions and the temperature graphs. All participants underwent two phases of training. After each phase, performance was assessed. Once again, a full ERE (disordinal interaction) was found. In the Phase 1 assessment (i.e., when students were low in knowledge), students in the practice condition performed better. In the Phase 2 assessment (after the students were higher in knowledge), students in the imagination condition performed better. In sum, the expertise reversal effect also applies to imagination. When individuals know what to do based on prior experience, imagination can be an effective form of practice. In the absence of sufficient experience/prior knowledge, imagination appears to hinder performance.

Finally, an interesting finding (which we dub the modality effect) pertinent to EREs has been obtained by manipulating presentation modalities and examining how split-attention and redundancy effects are impacted. Kalyuga, Chandler, and Sweller (1999) examined the implications of Baddeley's working memory model (1992) for EREs. Baddeley's model postulates two (largely) independent subcomponents of working memory: the phonological loop (PL) and visuo-spatial sketchpad (VS). The former processes auditory information, the latter visual.

In the most relevant comparison, Kalyuga et al. (1999) found that the split-attention effect could be overcome by presenting information in visual and auditory format. Phrased differently, if a low prior knowledge participant must integrate sources of non-redundant information, that integration is more easily accomplished if one source of information is visual and the other auditory than if both are presented visually. Per Baddeley's model, this is because the VS and the PL are not in competition for mental resources.

***Summarizing the expertise reversal effect literature.*** To our knowledge, there is no quantitative meta-analysis of ERE findings. There have been attempts to narratively summarize this research (Kalyuga, 2007; Sweller, van Merriënboer, & Paas, 1998), and obviously we have resorted to the same approach in our discussion of the various ERE effects.

One way to gain a feel for the kinds of domains that have been examined is to simply list them (see Table 1). Kalyuga (2007) provides a helpful chart listing ERE findings with respect to references, experimental conditions, sample sizes, and effect sizes. We chose to replicate the general structure of that chart, but focusing upon the knowledge domain and participant samples used. All of the articles cited in our review are included in this chart, and the references in the chart overlap to a large extent with those of Kalyuga (2007). We are aware that there are references listed in the chart which were not discussed in our narrative summary above. In such cases, they were conceptual replications or extensions of findings we did discuss. We also list the manner in which prior knowledge was operationally defined, as well as whether that operational definition should be classified as ‘induced’ or ‘measured’.

Table 1.  
*Summary Table of Expertise Reversal Effects Literature*

<b>Reference</b>	<b>Sample</b>	<b>Effect</b>	<b>Domain</b>	<b>Prior Knowledge</b>
Kalyuga, Chandler, & Sweller (1998)	Exp 1: 26 1 <sup>st</sup> -year mechanical apprentices	Split attention effect	Engineering	<u>Measured</u> by ensuring minimal prior experience through use of 1 <sup>st</sup> -year apprentices
	Exp 2: 33 1 <sup>st</sup> -year mechanical apprentices	Split attention and redundancy effects	Engineering	<u>Induced</u> by measuring performance across blocks of trials
	Exp 3: 33 1 <sup>st</sup> -year mechanical apprentices	Split-attention and redundancy effects	Engineering	<u>Measured</u> by ensuring highly experienced through use of same sample as Experiment 2
Yeung (1999)	Exp 1: 134 5 <sup>th</sup> grade English as Second Language (ESL) students from Hong Kong	Spit-attention and redundancy effects	Comprehension of written English	<u>Measured</u> by ensuring minimal prior knowledge through use of 5 <sup>th</sup> grade ESL students
	Exp 2: 126 8 <sup>th</sup> -grade ESL students from Hong Kong	Split-attention and redundancy effects	Comprehension of written English	<u>Measured</u> by ensuring higher prior knowledge (respectively to Exp 1) through use of 8 <sup>th</sup> -grade ESL students
	Exp 3: 25 1 <sup>st</sup> -year ESL university students from Hong Kong	Split-attention vs. redundancy effects	Comprehension of written English	<u>Measured</u> by ensuring higher prior knowledge (respectively to Exp 1 & 2) through use of 1 <sup>st</sup> year ESL students
McNamara, Kintsch, Songer, & Kintsch (1996)	Exp 2: 56 7 <sup>th</sup> through 10 <sup>th</sup> graders	Redundancy effect	Heart disease	<u>Measured</u> by median split of pre-test on heart functions

Table 1. (continued)

Kalyuga, Chandler, & Sweller (1999)	Exp 1: 34 1 <sup>st</sup> -year mechanical apprentices	Split-attention, redundancy, and modality effects	Engineering	<u>Measured</u> by ensuring minimal prior experience through use of 1 <sup>st</sup> -year apprentices
	Exp 2: 16 1 <sup>st</sup> -year mechanical apprentices	Split-attention and redundancy effects	Engineering	<u>Measured</u> by ensuring minimal prior experience through use of 1 <sup>st</sup> -year apprentices
Mousavi, Low, & Sweller (1995)	Exp 1 – Exp 5: 8 <sup>th</sup> grade students	Split-attention effect	Geometry	<u>Measured</u> by selecting students with highest math grades
Tindall-Ford, Chandler, & Sweller (1997)	Exp 1: 30 1 <sup>st</sup> -year trade apprentices	Split-attention and redundancy effects	Engineering	<u>Induced</u> by measuring performance across blocks of trials
	Exp 2: 22 1 <sup>st</sup> -year trade apprentices	Split-attention and redundancy effects	Engineering	<u>Induced</u> by measuring performance across blocks of trials
	Exp 3: 24 1 <sup>st</sup> -year trade apprentices	Split-attention and redundancy effects	Engineering	<u>Induced</u> by measuring performance across blocks of trials
Tuovinen & Sweller (1999)	32 Diploma of Education students	Worked examples vs. exploratory learning	Using a database	<u>Measured</u> by having students rate frequency of database usage
Kalyuga, Chandler, & Sweller (2001)	Exp 1: 17 1 <sup>st</sup> -year trade apprentices	Worked examples vs. exploratory learning	Engineering	<u>Measured</u> by ensuring low prior knowledge through use of 1 <sup>st</sup> -year trade apprentices
	Exp 2: 17 1 <sup>st</sup> -year trade apprentices	Worked examples vs. exploratory learning	Engineering	<u>Induced</u> by measuring performance across blocks of trials
Kalyuga, Chandler, Tuovinen & Sweller, (2001)	Exp 1: 24 1 <sup>st</sup> -year trade apprentices	Worked examples vs. problem solving	Engineering	<u>Induced</u> by measuring performance across blocks of trials
	Exp 2: 24 1 <sup>st</sup> -year trade apprentices	Worked examples vs. problem solving	Engineering	<u>Induced</u> by measuring performance across blocks of trials
Clark, Ayres, & Sweller (2005)	Exp 1: 24 9 <sup>th</sup> -graders	Technology effect	Spreadsheets and algebra	<u>Measured</u> by having students rate frequency of spreadsheet usage

Table 1. (continued)

Leahy & Sweller (2005)	Exp 1: 60 4 <sup>th</sup> grade students	Imagination effect	Using train time tables	<u>Induced</u> by measuring performance across blocks of trials
	Exp 2: 60 5 <sup>th</sup> grade students	Imagination effect	Use of temperature graphs	<u>Induced</u> by measuring performance across blocks of trials
Leahy & Sweller (2007)	30 3 <sup>rd</sup> grade students	Imagination effect	Bus time table	<u>Induced</u> by measuring performance across blocks of trials

*Note.* Kalyuga, Chandler, and Sweller (2001) examined both the sensory modality effect (Experiment 1) and split attention/redundancy effects (Experiment 2).

From Table 1, it is clear that the domains tackled are very linear, structured ones. The most commonly examined domain, for instance, is engineering. How replicable these findings are in less structured domains remains to be seen. However, it should also be noted that the age of the samples ranged from third graders to young adults, which indicates that these effects are not limited to specific age groups. This table also makes clear why we focused on certain effects (i.e., split-attention, redundancy, and worked examples) while briefly summarizing others (technology, imagination, and sensory modality). Of the 12 references cited, 6 examined split attention and/or redundancy effects, 3 examined worked examples, and only 4 examined the ‘emerging’ effects of technology, imagination, and sensory modality. (See footnote to Table 2 for why these numbers exceed the total number of references.) It should also be appreciated that the sensory modality effect is actually an ‘effect of effects’—a moderating variable that impacts how the split-attention effect is expressed.

**Ordinal ATI.** At the beginning of the ERE section, we stressed that not all ATI (whether involving prior knowledge or some other aptitude) are disordinal interactions. This point can be appreciated by making clear that prior knowledge (expertise) will often interact with structure, but it is not necessarily the case that a *disordinal* interaction will result. Put differently, we could say that while we might see an expertise effect, we might not see an expertise *reversal* effect. Whether or not such a reversal is seen is dependent on the range of prior knowledge present in the participant sample. The point of the experimental literature on EREs is to find that inflection point at which a reversal can occur. And finding such an inflection point is not always easy.

We can illustrate this by revisiting one of the studies cited in Table 1 above, namely that of Kalyuga, Chandler, and Sweller (1998). That article described a set of three experiments which sought to establish an ERE in which low prior knowledge individuals performed better with an integrated rather than separated display, and vice versa as expertise (prior knowledge) was gained. What the authors found in the second experiment, however, was hints of an as yet-to-be realized reversal.

Another way of seeing this point is to ponder the findings of Pascarella (1978). Pascarella found a disordinal interaction between prior mathematics knowledge and degree of structure. Although a disordinal interaction (ERE) was obtained, it only held for a relatively few

individuals who had obtained a very high level of prior knowledge. If those few individuals had not been present, then the reversal would not have been present. Rather, the interaction would have been ordinal in nature. A similar pattern was found by Ross and Rakow (1981). Although a disordinal interaction was found, it was driven by individuals whose prior knowledge scores exceeded those of 97% of the participant sample. As the authors put it, “the practical meaning of that result can be questioned” (p. 750). There is thus support for our contention that the range of prior knowledge present in a given sample (or population) is a factor affecting the nature of any ATI.

***ATI with military subject matter and populations.*** We now present ATI found in research using military subject matter domains. Both ordinal and disordinal interactions were found in these studies. The findings demonstrate that despite most ATI research being conducted in “laboratory settings,” individual differences in aptitude are relevant in the applied, military context as well and bear directly on deciding the most effective mode(s) of military instruction.

*Military subject matter with a non-military adult sample.* Work by Shute (1992; 1993) represents a systematic way of examining ATI in accordance with the second-generation ATI research cited by Pellegrino et al. (1999). The instruction involved technical military topics, but the participants were from a non-military adult population. There were methodological similarities in the two experiments. Individual differences were assessed via computer, and the training was computer-based. The experiments required a total of seven days, and each had large sample sizes (*ns* of 282 and 178). The large sample sizes distinguishes this body of work from much other ATI research. The subject matter was technical: Ohm’s law and basics of flight engineering (determining whether factors precluded or warranted a safe plane flight). In the 1992 work, associative learning (AL) was the cognitive aptitude of interest, while in the 1993 work, working memory (WM) and general knowledge (GK) were the two aptitudes of interest. In addition, both efforts included multiple criterion measures, based on cognitive learning theory, for which differences were expected as a function of the aptitude-treatment combinations. Shute pointed to the advantages of matching student aptitude to the learning environment. Lastly, Shute examined the cost-benefits and feasibility associated with tailored according to the recommended decision-rules. She concluded that costs were low as the computer time to test aptitudes was short and the computer algorithms underlying the treatment conditions were easy to change.

In the 1992 research, the aptitude was AL, and two treatments were compared: rule-application where learners were given feedback on the principle involved in solving problems and then students applied the rule, and rule-induction where learners were given general guidance regarding the relevant variables in the problem to enable them to generate their own interpretation of the underlying principles. Declarative knowledge and procedural skills were assessed. For declarative knowledge, there was a disordinal interaction, with high AL learners scoring better with rule-induction and low AL learners better with rule-application. For procedural skills there was an ordinal interaction, with high AL individuals scoring better in the rule application treatment, while for those with low AL skills there was little difference between the two treatments. These interactions led to decision-rules for the instructional objectives. When the objective is declarative knowledge, rule-induction should be used for high AL

learners, but rule-application should be used for low AL learners. However, when the objective is procedural skills, then rule-application should be used for all.

In the 1993 research, Shute used tests of WM and GK as aptitude measures. Two treatments that varied the amount of practice were compared: constrained practice where learners had 3 problems per problem set and extended practice where learners had 12 problems per set. Three criterion tests were used: knowledge and skills related to basic graphs, complex graphs, and flight engineering. Four aptitude profiles were generated – all possible high-low combinations on WM and GK. The results showed that individuals with a low-WM/high-GK profile benefited from the constrained practice environment, but that the extended practice environment worked best for the other learners.

*Military subject matter with Soldiers.* Three research efforts are cited here. In each case, aptitude treatment interactions were found, although that was not necessarily the original focus of the research.

With the first example, using an initial military training sample of Soldiers and map reading tasks, differences were expected as a function of level of knowledge (Wampler, Bink, & Cage, 2011). An ordinal interaction was found on retention scores. Specifically, the retention scores of low-performing Soldiers were significantly improved by hands-on practice with supplementary materials but retention did not improve when no materials were used. However, there was no impact on the retention scores of high-performing Soldiers in their map reading skills, regardless of training condition. This finding is consistent with other ATI research regarding the importance of instructional support for those with low prior knowledge.

The second example (Dyer et al., 2005) also supports the prior ATI research showing the importance of structure/instructional support for individuals with less knowledge and experience. The major purpose was to determine the most effective means of training digital map skills; the five instructional approaches were described earlier in this review. Although there was no formal measurement of prior knowledge in this case, the two groups compared, new Soldiers in their initial training (Infantry One Station Unit Training [OSUT]) and newly commissioned officers in their initial officer training (Infantry Officer Basic Course [IOBC]), differed in military background and experience. For the two conditions with the highest degree of structure (exercises were required in both), there was no difference between the two groups of Soldiers. But for the other three conditions, which had less instructional support, the two groups differed, with the least experienced Soldiers performing the poorest. The greatest difference occurred for the exploratory condition, which had the least instructional support.

The last example is rather unique as it deals with a very specific topic – unaided night vision (Dyer, Gaillard, McClure, & Osborne, 1995). Two instructional conditions were compared. One was an unaided night vision program given in the dark that used specially prepared slides to illustrate night phenomena such as loss of color vision, night blind spot, off-center vision, lessened visual acuity, the autokinetic illusion, effect of strobe lights, and how to protect night vision. Soldiers in this condition were exposed to slides that demonstrated these phenomena, and also saw word slides that described these night vision phenomena while an instructor talked about the information on the slides and the visual phenomena. The other

condition was simply a text version of the information on the slides, similar to information in Army Field Manuals. The aptitude measure was the General Technical (GT) score (reflecting verbal and mathematic abilities) from the Armed Services Vocational Aptitude Battery. The expectation was that the slide program would be the most effective for all Soldiers. However, a significant disordinal interaction was found. GT score was linearly related to scores for those in the text condition, but curvilinearly related to scores for those in the program condition. Those with the lowest GT scores did better with the program than the text. Those with the highest GT scores did better with the text than the program. Based on Soldier comments, the hypothesized reason for the lower performance by those with high GT scores with the program was that these Soldiers read the slides, and found it hard to overcome the habitual use of foveal vision (cones) and shift to off-center vision (rods) techniques necessary to read the slides, which were set at 20/50 visual acuity level typical of night vision. On the other hand, those with low GT scores indicated they simply listened to the instructor while observing the phenomena and did not focus intently on trying to read the word slides. The results show that when developing instructional materials and techniques, it is critical to fully analyze the cognitive progresses likely to be required for task performance.

**Explanation.** So far, we have been treating ATI as if they were brute facts—facts without any explanation. However, there is a proposed explanation for these phenomena. To that we now turn.

**Cognitive load theory.** One explanation of prior knowledge/structure involves the concept of cognitive load. Cognitive load theory (CLT) attempts to derive instructional design guidance from features of the human memory system (for a complete discussion of the postulates of CLT, see Sweller, van Merriënboer, & Paas, 1998). In brief, CLT is concerned with demands placed upon an individual's working memory. When an individual possesses little prior knowledge of a domain, cognitive load is high because working memory is heavily taxed. Therefore, novices benefit from treatments which minimize extrinsic (i.e., irrelevant) memory load and allow the learners to concentrate on the intrinsic (i.e., salient-determined by the nature of the content, not the instructional design) memory load. As domain knowledge increases, cognitive load decreases. Thus high prior knowledge individuals do not require, and are sometimes impeded by, highly structured treatments.

Understanding why load on working memory decreases with domain familiarity requires a brief discussion regarding the interplay between working memory and long-term memory (Cowan, 1988). Working memory may be thought of as that which stores the contents of current, conscious awareness, while long-term memory contains information to which we have access, but of which we are not currently aware. Although working memory appears to have a quite limited capacity in some contexts (Miller, 1956), in other contexts working memory appears to be quite capacious—e.g., experts in their areas of specialization can often handle vast amounts of complex information. Understanding these apparently contradictory phenomena has led to the realization that the working memory and long-term memory systems interact much more than was previously thought—so much so that some researchers (Ericsson & Kintsch, 1995) have proposed amending traditional working memory models to include a feature explaining the interplay between working memory and long-term memory.

As domain familiarity within a field increases, domain knowledge stored in long-term memory increasingly becomes organized into “schemas.” Schemas have been variously defined as the extraction of general, essential information from individual instances (Chen & Mo, 2004), as abstract categories that individual instances instantiate in different ways (Gick & Holyoak, 1983), or as constructs allowing problem solvers to group problems into sets (Cooper & Sweller, 1987). However exactly one wishes to define schema, the general consensus is that schema in long term memory help in the comprehension (Bransford & Johnson, 1972, as cited in Eysenck & Keane, 2005) and recall (Chi, Glaser, & Rees, 1982; Lambiotte & Dansereau, 1992) of new but related material. In other words, increasing knowledge within a domain leads to the development of schemas in long-term memory. This statement is well supported, as these effects have been found in chess players (De Groot, 1966), professional musicians (Halpern & Bower, 1982; Kalakoski, 2008), and medical students (Arocha & Patel, 1995). Those schemas aid in more efficient “chunking” of information in short-term memory. This reduces the load on working memory, which in turn means that the learner now requires less support (i.e., less structured) treatment. Finally, this allows the learner to grapple with the relevant (intrinsic, inherent, or content-driven) cognitive load imposed by the material.

The goal of CLT is to find the ‘sweet spot’—to provide structured treatments so that extrinsic (irrelevant) cognitive load on working memory is minimized during the novice stages. As domain knowledge is gained and long-term memory schemas develop, treatment conditions become less structured and more and more of the burden for learning is placed on the learner.

***Cognitive load theory and EREs.*** There are essentially two lines of evidence for CLT. The first line of evidence arises from the fact that the various EREs which exist are predicted by CLT. The second line of evidence involves measures of cognitive load, to which we will turn shortly. First, however, we will examine just two EREs (the redundancy and split attention effects) and outline how CLT attempts to explain them.

In the redundancy and split-attention literature, the idea revolves around what information participants will already have stored in long term memory schemas. When exposed to information which is presented both in informational statements and diagrammatically, experts are at a disadvantage. The overlapping information is redundant, as their long term memory schemas already contain the ‘know how’ of translating the diagram into propositions. Thus, part of working memory resources must be devoted to filtering out irrelevant information. For novices lacking such schema, having the additional information available is helpful. As the chunks in the novice working memory systems can contain only limited amounts of information, being spared the process of translating the diagrams into sentence form presumably relieves an overtaxed working memory system.

The second line of evidence supporting a cognitive load theory interpretation of expertise reversal effects comes from what Kalyuga, Chandler, and Sweller (1998) have called ‘subjective ratings of mental effort.’ Typically, these ratings are simple Likert scales in which participants are asked to rate how easy or difficult some material was to understand (1 = extremely easy to 7 = extremely difficult). In general, the pattern of subjective ratings of mental effort maps well onto performance predictions generated by CLT. For example, turn again to the redundancy and split-attention effects. CLT predicts that novices should exhibit higher subjective load with

separate displays versus single displays, and the opposite should be true for experts. Both effects have been found (Kalyuga et al., 1998). CLT predictions regarding subjective load have also been supported in regards to the modality effect (Kalyuga, Chandler, & Sweller, 1999), the worked example effect (Tuovinen & Sweller, 1999), and the technology effect (Clark, Ayres, & Sweller, 2005).

***Cognitive load theory and tailored training.*** CLT indicates that various features of the domain or tasks to be trained (i.e., element interactivity) and of the training population (degree of prior knowledge) need to be considered. As Sweller, van Merriënboer, and Paas (1998) note, the joint effects of task and student characteristics are sometimes salient. Although some types of material possess higher intrinsic cognitive load (due, e.g., to element interactivity) than others, the amount they impose upon a given individual will vary in accord with the individual's prior knowledge of that domain. This suggests that training developers and/or trainers/instructors should attempt to measure the intrinsic cognitive load of a domain (see Sweller, 1994), take into account the domain knowledge of their training population, and design instructional procedures and materials accordingly.

## **Scaffolding**

The central message of the ERE literature is that low prior knowledge individuals require more highly structured environments than high prior knowledge individuals. Although the ERE literature is primarily a product of the last decade, most ERE treatment manipulations fall under the concept of scaffolding (Renkl, Atkinson, Maier, & Staley, 2002; Snow, 1992), an older concept within educational research. Scaffolding has been mentioned in previous sections of this paper, and is a theme that runs throughout the literature. Here we discuss the concept in more detail.

Scaffolding can take several different forms, including leveraging social interactions between tutor and tutee (Van Lehn, 2011) and computer-based learning (Hogan & Pressley, 1997). In broad strokes, scaffolding involves an activity or task that is appropriate for the individual's aptitude level (Applebee & Langer, 1983; Gallimore & Tharp, 1983; Hmelo-Silver, Duncan, & Chinn, 2007). By providing support for learning, new information is integrated with existing knowledge because the task provides a context and motive for integration (Brown, Collins, & Duguid, 1989; Johnson-Laird, 1995; Langer & Applebee, 1986; Wickens, 1987).

Usually scaffolding is applied individually and is gradually removed (or faded) as the learner gains familiarity with a task (Wood, Bruner, & Ross, 1976). Also as cited in the tutoring sections of this review, scaffolding is often adjusted or lessened. Tutors provide just enough information to enable students to complete tasks on their own or provide less support for students' errors when the learning environment presents significant benefits to students who can solve problems on their own. Thus, scaffolding involves learning increasingly difficult tasks coupled with diminishing support for those tasks. Extensive scaffolding also requires ongoing evaluation of the individual's performance. This evaluation may not only be used as feedback for the learner, but also to calibrate task difficulty and level of support.

Arguably, all of the prior ERE results can be tied to scaffolding. In the case of the redundancy and split-attention effects, the disordinal interaction is driven by the fact that failing to remove (fade) support as expertise is gained results in inferior performance. A similar argument can be made for the technology and modality effects.

One prominent scaffolding technique which arises from the ERE literature (specifically, the worked example effect) is backward fading. The gist of the worked example effect, recall, is that providing novices with worked examples provides sufficient scaffolding to enable learning. As expertise is gained, problem solving (i.e., placing responsibility for the successful execution of all steps for solution on the learner) becomes appropriate. Backward fading is simply a means of gradually moving individuals from the worked examples to problem solving.

For the fading approach to work, the knowledge domain must be *cumulative*. It implicitly assumes that there are cumulative relationships between the steps used to solve given problems. For this reason, fading approaches have largely been examined within mathematics or mathematically-related domains (Renkl, Atkinson, Maier, & Staley, 2002). Understanding the rationale underlying the second step in a problem-solving process, therefore, presumably requires understanding the rationale underlying the first step.

In *backward* fading—as might be expected—the complete worked example is shown first, then the last solution step is left blank for the learner to provide. Then, the last two steps must be provided by the learner. Finally, the learner is confronted with a practice problem, all the solution steps of which must be provided solely by the learner (Renkl et al., 2002; Shen & Tsai, 2009). Scaffolding has also proven effective with perceptual tasks (Salomon, 1974).

Viewing the instructional approaches from the expertise reversal literature through the lens of scaffolding accomplishes several things. First, it helps to make even clearer the rationale underlying cognitive load theory and expertise reversal effects. Secondly, it plausibly increases confidence in the findings from that literature, as it anchors them in a well-researched approach to instructional design. Third, the scaffolding approach fits well with the “crawl, walk, run” methodology of U.S Army training. Fourth, it helps sharpen what should be involved in tailoring training: identifying existing knowledge/skill levels, providing the appropriate amount and kind of support when needed, and removing that scaffolding when no longer required. But it doesn’t necessarily address the issue of initially identifying those who do not need scaffolding nor of how to adapt to different levels of structure needed once training has started and individuals progress at different rates.

It is plausible that hands-on tasks which have heavy cognitive and perceptual (i.e., visual) components would also benefit. However, as can be seen by perusing the literature results discussed thus far, systematic research on scaffolding with hands-on tasks is to our knowledge almost non-existent. A current research effort is underway examining how backward fading techniques might be used to train hands-on military tasks. At the moment, however, this hole in the literature indicates that military research in applying backward fading or other systematically scaffolded techniques is required.

## **Literature Review Summary**

In this section, we recap the major conclusions of the literature review and draw out some implications for tailoring training in U.S. Army institutional settings. In the final section of the paper, we discuss issues in transitioning research findings to Army settings and draw some recommendations for future research in tailoring Army institutional training. Before we begin the summary, however, recall that the purposes of this paper were to (1) examine the research literature and isolate the major areas of tailored training research, (2) determine which types of tailored training seem to be most effective and under what conditions, and (3) provide suggestions for future tailored training research with near-term applicability in Army settings. In the literature review we addressed the first purpose. Our goal in this summary is to draw together some of the threads of the literature review and more specifically address the second purpose.

### **Ability Grouping**

The central message of the ability grouping literature is that effects are driven by instructional factors rather than institutional or social ones. In addition, the types of ability grouping which demonstrate the largest effect sizes are also those in which the most tailoring occurs. In other words, simply grouping individuals of similar aptitude together in the absence of training is not likely to be very fruitful. Of the five ability grouping approaches, within-class, enriched, and accelerated appear to be the most relevant to Army courses. Within-class grouping would be appropriate with a wide range of Soldiers, while the enriched and accelerated are designed for high-ability Soldiers.

### **Learning in Small Groups**

In many ways, the applied research literature on the use of small groups in the public schools is not directly relevant to military classrooms. Although improving learning was one goal, the primary focus was often on developing cooperative skills, using well-defined subject-matter domains and often varying the distribution of ability within the group. In contrast, in military settings, small groups are frequent, and deviate from the lecture mode. Such groups typically work on open-ended type tasks where there is not necessarily a single best solution (Dyer, et al. 2011). Group composition is often based on the prior military experience of individuals, not ability. The best evidence for what makes small groups effective comes from intensive observations of the group learning process where individuals work on problems without teacher assistance. A common finding is that individuals who receive nonresponsive feedback from others in the group learn less than individuals who receive responsive feedback, and individuals who give explanations achieve more. Overall, it is clear that just breaking individuals into groups within a classroom does not guarantee effective adaptation to individual learner's strengths or weaknesses, as peer group members are not necessarily skilled in how to make the group an effective learning setting for each member. However, with relatively senior military personnel and research-based guidelines, instructors should be able to facilitate tailored training in small group settings.

## **Tutoring**

The tutoring literature shows that tutoring is effective, with most research conducted in reading and mathematics, and often in remedial contexts. Beyond these general findings, the primary question of interest is what tutors do that make them effective. Analysis of student-tutor dialogues indicates that the most effective tutors are experts in their domain, use pedagogical techniques that allow them to understand the student's point of view, use scaffolding techniques as opposed to didactic techniques, tailor their feedback depending on the type of student error rather than simply saying "yes" or "no," and use Socratic reasoning approaches. What the student does is also important. Students learn more when they contribute ideas and ask specific questions that help the tutor understand their knowledge and perspective of the subject matter. And when tutors are trained to ask open-ended questions of students rather than depending on giving extra information and explanations, students begin to talk more and the tutors become more interactive, less didactic and used more scaffolding prompts. Thus, not only does the tutoring literature show that tutoring is effective, there is a relatively good foundation of knowledge regarding what good tutors do, and therefore what techniques would be effective when military instructors provide one-on-one training. The challenge appears to be in quickly preparing these instructors to be good tutors.

## **Microadaptation**

Microadaptation is widely recognized as important, but is largely neglected as a systematic research topic (Corno, 2008; Nuthall, 2004). Because microadaptation requires both extensive domain knowledge and practice teaching that domain, many teachers do not micro-adapt very effectively (Clark & Yinger, 1977). This suggests that instructors might not have enough time to develop effective microadaptation, given the relatively rapid turnover in Army course instructors. Ways to offset this might involve structured materials illustrating good pedagogy for incoming instructors or implementing some rigorous macro-adaptive process like repeated cycles of training/assessment, and validated materials varying in structure according to progress (i.e., domain knowledge) within a course.

## **Learning Styles**

There are several major problems with the learning styles literature, including insufficient validation of instruments (Curry, 1990), failure to establish measurement reliability (Coffield, Moseley, Hall, & Ecclestone, 2004), and failure to demonstrate replicable interactions between learning styles and learning conditions (Pashler et al., 2009). This leads to the conclusion that investigating tailored training efforts through the lens of learning styles is not a judicious use of resources.

## **Aptitude-Treatment Interactions**

Aptitude-treatment interactions (either ordinal or disordinal) provide the most systematic window into understanding interactions between salient individual differences (aptitudes) and instructional conditions (treatments). This is not to suggest that this approach is the be-all-end-

all of tailored training. Nonetheless, it does provide some clear recommendations regarding tailoring training.

First, this approach suggests that prior knowledge (via its relationship with the ‘upstream’ variables of experience and general mental ability) is the prime aptitude to focus on. Prior knowledge can be either general or domain-specific. Second, it suggests—at least in broad strokes—the kinds of conditions under which low and high prior knowledge individuals should be trained. Namely, low prior knowledge individuals should be provided with a high degree of structure or scaffolding, which is then removed as domain knowledge is gained. High prior knowledge individuals, once brought to a sufficiently high level of performance, should be provided with the opportunity to practice the skill to the point of automaticity—without interfering (previously beneficial) scaffolding.

Research shows that both disordinal and ordinal interactions occur, yet which will occur is not easily predicted. Therefore, empirical investigation is necessary to determine the exact nature of the interaction. Both types of interactions have implications for decision-rules regarding the conduct of tailoring training. If a disordinal interaction is the primary pattern, then different approaches are needed for students with different aptitudes, which makes the tailoring more complicated to execute. If an ordinal interaction occurs, the best approach for those with low aptitude measures may work well for those who with high aptitude or the opposite could be the case (reference the diagrams in Figure 1). If the instructional material is delivered via computer, it may be relatively easy to adapt to individual differences regardless of the nature of the interaction, but if the instruction is face-to-face then adaptation may present substantial logistical problems.

### **Issues in Applying Research Findings**

In the next section, we address the third objective of the paper: to provide suggestions for tailored training research with near-term applicability in Army settings. However, before we do so, we present in this section basic characteristics of Army institutional training based on our prior training research and experience. These points should not be unfamiliar to the reader, but provide a context for this section. In addition, these points are important because they have implications for future research, specifically on the generalizability of academic research findings to Army settings.

First, the Soldier population is an adult population whose training is preparation for a future job or duty position. The ability grouping research, by contrast, was conducted with elementary and middle school children. Further, many if not most of the participants used in the academic research settings were not preparing for a duty position or career, so there are definite motivational differences. Second, Soldiers within most classes have diverse backgrounds making it difficult to assume they begin a class with the same prior knowledge. This variability in prior knowledge is plausibly wider than the variability in prior reading knowledge for young children entering first or second grade. As prior knowledge plays such a central role in predicting performance, this is not an insignificant consideration. Third, the subject matter covered in institutional training reflects a spectrum of cognitive, procedural, hands-on, analytic,

technical, leader, and other skills. Fourth, training media and methods vary greatly, as might be expected given the diversity of subject matter and training requirements. Thus you will find small and large classes, one-on-one training, face-to-face instruction as well as distance learning, problem-solving sessions, hands-on training, field training, and use of training devices and simulations. Fifth, for some courses, instructors are entirely military who are assigned a training position for about a two- to three-year period; in other courses, there can be both military and civilian instructors. Sixth, students in military courses are highly motivated as the skills and knowledge they acquire are directly relevant to their profession where that is not the case with most research settings. Even a cursory review of these general points by the reader should indicate that there are major differences between the tailored training research reviewed and military instructional settings. Regardless, we believe there are research findings and conclusions that apply to military settings, which show potential for application, as well as present challenges to tailoring training.

### **Subject Matter Domains**

The majority of domains in which ATI have been found, as well as where other tailored techniques have been applied, are highly structured and cumulative (e.g., engineering, geometry, reading). Such domains characterize some (but hardly all) of Army training. It is therefore unclear how suitable more ill-defined domains would be for such tailoring. Conversely, cumulative, highly-structured domains should allow instructors to more easily track skill acquisition. A course in engineering might, for example, feature a series of modules which are very sequential, with later modules building upon the skills and knowledge gained in previous modules. This allows instructors to (a) estimate the statistical relationships among various criteria in the course (b) insert prior knowledge checks at multiple points and (c) make informed adjustments to materials, presentation rate, or other aspects of instructional design to accommodate differences in skill level.

While such advantages should arise in such domains, it nonetheless remains difficult to give practical advice on how to specifically instantiate such processes. This is so for the simple reason that ATI and other tailoring studies are often short experiments rather than multi-week courses, so extrapolating the appropriate tailoring approaches to a lengthy Army course is somewhat hazardous. Instructors may have difficulty maintaining a specific training approach (high instructional support vs. low instructional support) to training over an extended period of time. In addition, adjustments may be necessary as the individual difference profile of the Soldier-students (knowledge and expertise gained) can change with training.

### **Measuring Prior Knowledge**

Given the conclusion that prior knowledge is the most direct and hence the most powerful predictor of domain performance, we recommend that prior knowledge be used as the primary (perhaps only) aptitude. Prior knowledge has two advantages above and beyond its predictive power. First, it is 'tractable' in a way that measures of cognitive styles, personality, and general mental ability may not be. Second, measures of prior knowledge will likely have more 'face validity' than those other types of measures. A military instructor will be demotivated to use a measure that does not have a prima facie link to the course content, and Soldier-students might

also be demotivated when completing such measures. Third, such assessments can be used to identify misconceptions and naïve understandings on part of the Soldier-student, which have been shown to inhibit learning (Dochy, Segers, & Buehl, 1999), and provide indications to the instructor that efforts must be instigated to counter misconceptions.

For prior knowledge to predict performance accurately, however, it must be objectively assessed, not subjectively estimated (Shapiro, 2004). Further, where possible, the prior knowledge measure should be narrowly targeted, what Shapiro called topic knowledge. In other words, asking very general questions about a domain may fail to reveal meaningful differences in prior knowledge. However, when a Soldier's knowledge in a specific domain is limited, assessing general or domain-level knowledge may prove useful (Shapiro, 2004; Shute, 1993). Another factor to consider in assessing individual differences is to base such tests on differences between experts and novices (Pellegrino et al., 1999).

Just as the measures of prior knowledge should be as objective as possible, so too should the course criteria. This might require re-thinking how course criteria are developed and implemented. In at least one case with which we are familiar, 'Go' status was differentiated from 'No Go' on the basis of a simple median-split on a course criterion....not on how that outcome related to other course criteria. Shute's work (1992, 1993) illustrates how criterion tests can reflect both declarative knowledge and procedural skills, both important in military training.

One should also consider what is meant by 'prior knowledge.' The term is used somewhat ambiguously within the research literature itself. In some instances, prior knowledge seems to mean something like general domain knowledge, not necessarily knowledge that will be directly or explicitly tapped on some criterion (e.g., the ability grouping literature often uses a general measure of domain achievement) or something akin to prerequisite knowledge (Schaefer, Blankenbeckler, & Brogdon, 2011). In other cases, prior knowledge is tied clearly and obviously to criterion performance (Schmidt & Hunter, 1986).

In sum, while prior knowledge holds promise for Army tailored training, there are nonetheless several caveats regarding its use. We turn next to some of those caveats.

**Domains, measures, and changes in content.** One difference between the tailored training research settings and Army institutional training settings involves what kinds of measures are available, and how often such measures must be (re)validated. For example, the domains examined in the research literature are mathematics, reading, and engineering. For most of those domains, there are widely available, standardized achievement measures. Another difference is that the basics of learning mathematics, reading, and engineering probably do not change all that much—meaning that the achievement measures do not have to be changed very often or very much.

Neither is necessarily true of Army courses. There are probably very few widely accepted, standardized measures of achievement available to assess appropriate domain knowledge for Army courses. In addition, changes in doctrine or technology might require a whole-sale re-engineering of achievement measures. This implies a large and recurring resource investment.

Of course, if large-scale changes in achievement measures and course content are necessitated, it will also become necessary to re-construct and/or re-validate internal course criteria. However, there are several potential payoffs to so examining course criteria. First, finely tuned tailoring of training requires ongoing ('in stride') assessments of performance. A good example of where continuous measures can be helpful in tailored training was cited in prior research on marksmanship (Human Resources Research Office, 1959). Successive tests permit "early detection of exceptionally unskilled and skilled trainees. Additional instructor attention can then be given to the unskilled, while the skilled group is available as assistant coaches where required" (p. 6).

Even if multiple knowledge checks are not embedded between existing course criteria, understanding the statistical relationships among course criteria can be helpful. For example, knowing how prior poor performance on a criterion measuring performance on a first block of instruction is related to performance on a criterion on a second block of instruction would help instructors zero in on individuals who need assistance. Alternatively, knowing such relationships might also aid instructors in identifying individuals who might benefit from advanced training. Such relationships might also reveal changing predictor/criterion relationships as training progresses.

Knowing how criteria are related to one another might also help address a shortcoming of prior knowledge as a variable. Namely, administering prior knowledge measures to incoming Soldier-students is fruitful only if there are meaningful variations in prior knowledge among the incoming Soldier-students. However, if an early course criterion has known (i.e., empirically validated) relationships with later course criteria, then those early criteria can serve, in effect, as prior knowledge predictors for later course performance.

It must also be understood that not only would the development and validation of prior knowledge measures take substantial time and resources, but administering them at strategic points within a course would also take time. One potential time-saving approach involves the use of 'partial tests' (see Kalyuga & Sweller, 2004; Kalyuga, 2006a, 2006b, 2008). The core idea behind a partial test involves teaching a task which is highly structured and sequential in nature. Consider, for example, how to conduct a t-test in a statistical software package. When the user consults the software help file, there will be, say, seven steps on how to access the t-test function. One could estimate how much a student knows about that function by providing the first two steps and then asking them to indicate what the third step is. The relationship between the correctness of their response and a more detailed measure of performance could then be demonstrated. Such relationships have been found to be quite robust, and have been used to generate known EREs (again, see Kalyuga & Sweller, 2004). While the application of this testing approach to non-sequential tasks has not yet been adequately evaluated, this is one way of reducing time cost. Giving tests via computer is another way of reducing time (Shute, 1992, 1993).

**Deployment and operational tempo.** Finally, it must be understood that deployments and operational tempo have implications for measurements of prior knowledge. For example, in the research literature participants might involve 7<sup>th</sup> graders versus 9<sup>th</sup> graders, all of whom just

finished a given math course (or courses). There is little reason to suppose that there are substantial differences between the groups regarding the recency of education or exposure to information. This is most definitely not the case with the Army.

For example, a recent research effort looked at the relationship between various course predictors and exam performance in the Engineer Captains Career Course (Schaefer, Blankenbeckler, & Lipinski, 2011). The course was composed of officers with and without prior noncommissioned officer (NCO) experience. Both groups were equivalent on the prior knowledge measure, but prior knowledge was much more predictive of exam performance for those officers without prior NCO experience ( $r = .53$ ) than those with NCO experience ( $r = .11$ ). While this result cries out for replication, there are a few plausible hypotheses for this result. First is the fact that officers with prior NCO experience might have simply ‘absorbed’ enough of the lingo to be able to answer multiple-choice items as well as officers without NCO experience. When that prior knowledge base began to be built upon, however, perhaps those officers without NCO experience (who had probably been more recently exposed to those concepts) were better able to make conceptual links between their prior knowledge and the course content. More comprehension-type questions might have revealed meaningful differences whereas the multiple choice items did not (see McNamara, Kintsch, Songer, & Kintsch, 1996).

If such population differences are suspected, then prior knowledge measures composed of ‘deeper’ comprehension type questions might be appropriate. Alternatively, if performance is largely hands-on, then a hands-on predictor will probably be more predictive than a paper and pencil predictor (Schaefer, Blankenbeckler, & Brogdon, 2011). Of course, equipment availability may make the use of paper-and-pencil predictors preferable.

### **Instructor Preparation**

Another distinction between the academic literature and Army settings involves teachers/researchers vs. instructors. In the case of the research literature, experiments are usually administered by researchers and ability grouping by teachers. For many researchers and teachers, research and teaching are lifelong professions. They have had opportunity to develop a breadth and depth of knowledge that allows them to hone their skills in diagnosing and remedying miscomprehensions (in the case of teachers) or weaknesses of design (in the case of researchers).

Again, this is not necessarily true of the Army. Many Army course instructors are selected on the basis of necessity, not individual choice. In addition, many instructors will serve for only a few years in that position and then be given another duty assignment. This means that the instructor will likely not develop a broad and deep expertise, both within his/her field and in teaching that field, and will not benefit from the domain and pedagogical expertise of his/her predecessor.

In addition, even assuming that researchers have helped develop and evaluate prior knowledge measures and course criteria measures, it is ultimately the instructor who will have to interpret the resulting scores to assign individuals to various tailored training conditions. This is probably more complicated than it at first seems. For example, one under-appreciated aspect of

the ERE literature is that *where* the inflection point (defined as the point at which the relationship between knowledge and treatment changes) occurs is not known a priori. The researchers are free to change task complexity, or sample from a more diversely-experienced body of individuals, in attempts to locate that point. This is not the case for instructors.

Consider also the other means of tailoring applicable to the military. Based on the tutoring literature, the ways instructors can effectively address individual differences during one-on-one training sessions are known. In contrast, there is less empirical evidence on how to use small groups to facilitate tailoring beyond that of systematic assignment of individuals to groups which simulate future duty assignments which represents a very general adaptation to their career paths. Microadaptation techniques are likely to be used by experienced military instructors, but ways of preparing instructors with such techniques have not been systematically investigated.

Finally, military instructors have indicated that they are often unprepared to tailor systematically (Dyer et al., 2011), both because of lack of training and lack of available supplemental materials. This lack of preparation would be further compounded if changes in doctrine or technology required substantial overhaul of course content, criteria, and prior knowledge measures.

### **Recommendations for Future Research**

It seems to us that there are three broad areas of fruitful, near-term tailored training research opportunities in Army institutional settings: small groups, tutoring/microadaptation, and ATI. We address the first two in broad strokes, and spend some time detailing a possible approach in addressing the last. Interwoven throughout the remaining discussion are various factors related to feasibility of implementation and the various caveats outlined under the transitioning issues section. For all areas, we believe that the critical aptitude relevant to tailored training in the military should be prior knowledge, whether assessed at the start of a course, assessed for only critical parts of a course, or assessed continually throughout training.

#### **Small Groups**

Initially, small-group research in the Army should focus on an extensive and intensive examination of what typically happens in a variety of small-group military training settings. For example, whether individuals receive instruction and/or feedback appropriate to their learning status, how individuals are assigned to groups, and how the interaction among individuals varies with group composition and the training objectives. Given this information and any potential differences between what has proven effective in the research literature, the next phase would be to focus on how to enable instructors and/or their peers to become more effective in their instructional interactions with group members. The relative effectiveness of cooperative, collaborative, and intergroup competition within the military context could be compared. The limitations of small groups in providing individualized instruction, in contrast with tutoring and microadaptation approaches, should be examined. If the subject matter domains where small-groups exist typically involve planning and strategic thinking, or technical execution, then a core

part of the research should be on developing criterion measures that allow for tracking individual and group progress, and developing appropriate treatment conditions to allow tailoring.

### **Tutoring and Microadaptation**

The tutoring and microadaptation research overlap in the sense that effective tutors constantly microadapt, rather than use a precise script or pre-determined approach. When working with a single student, the tutor must react to the student's responses. Effective teachers in both settings are experts in their subject matter and have pedagogical expertise, and can thus draw upon their in-depth expertise in the instructional process. Within a military setting, it is most likely that tutoring (or one-on-one training) will occur when remedial training is needed. Thus the research approach would be to observe such remedial settings, document the instructor-student dialogue, obtain measures of student status and student progress, and determine what techniques and approaches were effective. Differences and commonalities in effective one-on-one military training settings could be compared with the existing literature. As much individual training in the military is hands-on, such research would also be an expansion to the subject matter domains typically investigated in the research literature. A follow-up to this would be research on how to best prepare new instructors for such settings.

### **Aptitude-Treatment Interactions**

If the goal is to determine critical aptitude-treatment interactions that exist within military settings and provide decision rules regarding aptitude-treatment combinations, then a two-phase approach is recommended. Phase I would be based on those settings and treatments to which the research findings are most likely to generalize. Phase II would incorporate the findings of Phase I and attempt to expand the applicability of tailored training methods to different types of subject areas and military settings.

Thus, Phase I would utilize military settings involving technical areas with cumulative subject matter, and where prior knowledge is expected to be the most critical individual difference variable. Stage 1 of Phase I might involve controlled, large sample, multi-day experiments like those of Shute (1992, 1993) to gain some solid information regarding the nature and extent of ATI that exist with military populations. Multiple Soldier populations and subject areas should be examined to determine if and how rank and experience impact the learning processes. Treatments would reflect variations in structure/instructional support. Sufficiently detailed criterion measures may have to be specifically constructed to reflect the depth and transfer of knowledge, but basic 'go/no go' baseline measures should also be included. Information would also be obtained on whether the treatments used benefited those with low aptitude, high aptitude, or both (i.e., were the interactions ordinal or disordinal?). This stage could also examine hands-on tasks or cognitive tasks with heavy hands-on components, as long as the other criteria (cumulative, technical) were met.

Stage 2 of Phase I would apply what was learned in the first stage, this time in actual technical courses. This would enable examining the feasibility of tailoring in ongoing classrooms, as well as examining the contribution of using continuing assessments as indicators of 'prior' knowledge. The approaches examined would address the instructor's primary need for

tailoring, whether remedial, advanced, or both (Jonasson & Grabowski, 1993). Supplemental criterion measures would again be included to assess the depth and extent of knowledge and skills gained. As before, the treatments would reflect degrees and types of structures.

Phase II would seek to extend the academic research findings and the results of Phase I to other types of domains. For example, while there is little to no ATI research dealing with decision making and planning skills, such skills are the foci of many military courses. The same Stage 1 (“basic”) to Stage 2 (“applied”) should be repeated with the non-technical courses. The program of research should yield generalizations regarding “what works” in the military. It should provide a solid instructional foundation for tailored training with guidelines on how to proceed with subject matter and Soldier populations not included in the research base.

### **Conclusion**

Regardless of what forms of tailoring are used in military settings, it is acknowledged that tailoring is not always warranted. A priori decisions regarding which phases of a course warrant strong attention to individual differences should be made, for example, perhaps phases of the course which Soldiers must master or where Soldiers often have difficulty. Based on the research literature, the critical individual different variable is that of prior knowledge regardless of the form of tailoring, which narrows the range of “aptitudes” considerably. But the techniques that best assess how much and what kind of prior knowledge is possessed by each Soldier and is directly relevant to tailoring must be addressed. In addition the tailoring techniques, whether in one-on-one situations, within small groups, or techniques designed for Soldiers with specific aptitudes must be thoroughly examined.

Within the tailored training literature, there are some indications of the direction in which near-term Army tailored training should go. There are, however, many empirical questions which must be answered before solid generalizations can be made from the tailored training academic research to Army institutional training settings. These generalizations must be made in light not only of evidence from academic and Army institutional tailored training research, but should keep in mind the joint considerations of implementation feasibility, military subject matter and training goals, and the fact that the training context is that of preparing individuals for the next phase of their military career.



## References

- Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, 47, 626-40.
- Alton-Lee, A. G., Nuthall, G. A., & Patrick, J. (1993). Reframing classroom research: A lesson from the private world of children. *Harvard Educational Review*, 63, 50-84.
- Anderson, L. W. (1985). A retrospective and prospective view of Bloom's 'learning for mastery.' In M. C. Wang & H. J. Walberg (Eds.), *Adapting instruction to individual differences* (pp. 254-268). Berkeley, CA: McCutchan
- Ansalone, G. (2001). Schooling, tracking, and inequality. *Journal of Children & Poverty*, 7, 33-47.
- Ansalone, G. (2009). Tracking, schooling, and the equality of educational opportunity. *Race, Gender, & Class*, 16, 174-184.
- Antil, L. R., Jenkins, J. R., Wayne, S. K., & Vadasy, P. F. (1998). Cooperative learning: Prevalence, conceptualizations, and the relation between research and practice. *American Educational Research Journal*, 35, 419-454.
- Applebee, A.N., & Langer, J.A. (1983). Instructional scaffolding: Reading and writing as natural language activities. *Language Arts*, 60, 168-175.
- Arocha, J. F., & Patel, V. L. (1995). Novice diagnostic reasoning in medicine: Accounting for clinical evidence. *Journal of the Learning Sciences*, 4, 355-384.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556-559.
- Berliner, D. C. (1985). How is adaptive education like water in Arizona? In M. C. Wang & H. J. Walberg (Eds.), *Adapting instruction to individual differences* (pp. 293-312). Berkeley, CA: McCutchan.
- Berliner, D. C. & Cahen, L. S. (1973). Trait-treatment interaction and learning. In F. N. Kerlinger (Ed.), *Review of Research in Education* (Vol.1, pp. 58-94). Itasca, IL: F. E. Peacock.
- Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.
- Borman, W. C., White, L. A., & Dorsey, D.W. (1995). Effects of rate task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology*, 80, 168-177.

- Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76*, 863-872.
- Bossert, S. T. (1989). Cooperative activities in the classroom. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol.15, pp. 225-250). Washington, DC: American Educational Research Association.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.
- Brody, L. E., & Mills, C. J. (2005). Talent search research: What have we learned? *High Ability Studies, 16*, 97-111.
- Brophy, J. E., & Good, T. L. (1970). Teachers' communications of differential expectations for childrens' classroom performance: Some behavior data. *Journal of Educational Psychology, 61*, 365-374.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32 – 42.
- Carbonaro, W. (2005). Tracking, students' effort, and academic achievement. *Sociology of Education, 78*, 27-49.
- Carroll, J. B. (1967). Instructional methods and individual differences. In R. M. Gagne (Ed.), *Learning and individual differences* (pp. 40-44). Columbus, OH: Charles Merrill.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.
- Chen, S. Y., & Paul, R. J. (2003). Editorial: Individual differences in web-based instruction— an overview. *British Journal of Educational Technology, 34*, 385-392.
- Chen, Z., & Mo, L. (2004). Schema induction in problem solving: A multidimensional analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 583-600.
- Chi, M.T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73-105.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Erlbaum.
- Chi, M. T. H., & Roy, M. (2010). How adaptive is an expert human tutor? In V. Aleven, J. Kay, & J. Mostow (Eds.), *International conference on intelligent tutoring systems* (pp. 401-412). Berlin: Springer-Verlag

- Chi, M.T. H., Roy, M., & Hausman, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32, 301-341.
- Chi, M. T. H., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, 22, 363-387.
- Chi, M. T. H., Siler, S. A., & Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471-533.
- Chorzempa, B. F., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. *Journal of Educational Psychology*, 98, 529-541.
- Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development*, 53, 15-24.
- Clark, C. M., & Yinger, R. J. (1977). Research on teacher thinking. *Curriculum Inquiry*, 7 (4), 279-304.
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning. A systematic and critical review*. London: Learning and Skills Research Centre.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64, 1-35
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79, 347-362.
- Corno, L. (1995). The principles of adaptive teaching. In A.C. Ornstein (Ed.), *Teaching: Theory into practice* (pp. 98-115). Boston: Allyn and Bacon.
- Corno, L. (2008). On teaching adaptively. *Educational Psychology*, 43, 161-173.
- Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., & Porteus, A. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 605-629). Washington, DC: American Educational Research Association.

- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104, 163-191.
- Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, 101, 779-789.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Curry, L. (1990). A critique of the research on learning styles. *Educational Leadership*, 48, 50-56.
- De Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In Kleinmuntz, B. (Ed.), *Problem Solving* (pp. 19-50). New York: Wiley.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1-19). Oxford: Elsevier.
- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evaluation of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.
- Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes on studies: The case of research on prior knowledge. *Review of Educational Research*, 69, 145-186.
- Durlach, P., & Ray, J. M., (2011). *Designing adaptive instructional environments: Insights from empirical evidence*. (Technical Report 1297). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA552167)
- Dyer, J. L. (2004). The measurement of individual and unit expertise. In: J. W. Ness, V. Tepe, & D. Ritzer (Eds), *Advances in human performance and cognitive engineering research, Vol. 5: The science and simulation of human performance* (pp. 11-124). Amsterdam: Elsevier.
- Dyer, J. L., Gaillard, K., McClure, N. R., & Osborn, S. M. (1995). *Evaluation of an unaided night vision instructional program for ground forces*. (Technical Report 1032). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA304276)

- Dyer, J. L., Singh, H., & Clark, T. L. (2005). *Computer-based training approaches for training interactive digital map displays*. (Research Report 1842). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA440171)
- Dyer, J. L., Wampler, R. L., & Blankenbeckler, P. N. (2011). *Tailored training in Army courses*. (Research Report 1950). Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA552439)
- Eash, M. J. (1961). Grouping: What have we learned? *Education Digest*, 27, 25-27.
- Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education*, 54, 151-162.
- Elbaum, B., Vaughn, S., Hughes, M. T. & Moody, S.W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92, 605-619.
- Eysenck, M. W., & Keane, M. T. (2005). *Cognitive psychology: A student's handbook*. East Sussex, U.K.: Psychology Press.
- Femlee, D., & Eder, D. (1983). Contextual effects in the classroom: The impact of ability groups on student attention. *Sociology of Education*, 56, 77-87.
- Fletcher, J. D. (1992). *Individualized systems of instruction* (IDA Document D-1190). Alexandria, VA: Institute for Defense Analyses (DTIC No. ADA255960)
- Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labeling, and spatial proximity. *Learning and Instruction*, 20, 216-224.
- Gallimore, R., & Tharp, R. G. (1983). *The regulatory functions of teacher questions: A microanalysis of reading comprehension lessons* (Technical Report 109). Honolulu: Kamehameha Educational Research Institute, The Kamehameha Schools.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, 59, 185-198.
- Gibbons, A. S., & Fairweather, P. G. (2000). Computer-based instruction. In S. Tobias & Fletcher, J. D. (Eds.), *Training and retraining: A handbook for business, industry, government, and the military* (pp. 410-442). New York: Macmillan.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39, 93-104.

- Good, T. L. (1988). Using the aptitude-treatment model to improve compensatory education: A call for more observational research. *Educational Psychologist, 23*, 357-363.
- Goska, R. E., & Ackerman, P. L. (1996). An aptitude-treatment interaction approach to transfer within training. *Journal of Educational Psychology, 88*, 249-259.
- Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents, 9*, 24-29.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104-137.
- Graesser, A. C., Sidney, S. K. & Cade, W. (2009). Instruction based on tutoring. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 408-426). New York: Routledge Press.
- Gully, S., & Chen, G. (2010). Individual differences, attribute-treatment interactions and training outcomes. In S. W. J. Kozlowski and E. Salas (Eds.), *Learning, training, and development in organizations* (pp. 3-64). New York: Routledge Taylor and Francis Group.
- Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in mathematics: A structural model. *Educational Psychology, 28*, 59-71.
- Hales, C. M. (1978). Accommodating individual differences. *Educational Leadership, 36*, 139-143.
- Hallam, S. (2002). Mixed up? The pros and cons of ability grouping. *Education Journal, 64*, 24-26.
- Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology, 95*, 31-50.
- Handelsmann, J., Egert-May, D., Beichner, R., Bruns, P., Change, A., & DeHaan, R. (2004). Scientific teaching. *Science, 304*, 521-522.
- Heathers, G. (1977). A working definition of individualized instruction. *Educational Leadership, 34*, 342-345.
- Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*, 3-15.

- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 99–107.
- Hogan, K. & Pressley, M. (1997). *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline.
- Human Resources Research Office. (1959). *TRAINFIRE V: Extension of research on TRAINFIRE I rifle marksmanship course* (HumRRO Research Memorandum). Fort Benning, GA: Human Resources Research Office, US Army Infantry Human Research Unit. (DTIC No. ADA479631)
- Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education*, 25, 343-358.
- Ireson, J., Hallam, S., Hack, S., Clark, H., & Plewis, I. (2002). Ability grouping in English secondary schools: Effects on attainment in English, mathematics, and science. *Educational Research and Evaluation*, 8, 299-318.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Johnson, D. W., & Johnson R. T. (2009). An educational psychology success story: Social Interdependence theory and cooperative learning. *Educational Researcher*, 38, 365-379.
- Johnson, D. W., Maruyama, G. M., Johnson, R., & Nelson, D. (1991). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89, 47-62.
- Johnson-Laird, P. N. (1995). Inference and mental models. In S.E. Newstead and J St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason*. Hove, UK: Erlbaum.
- Jonassen, D.H. & Grabowski, B.L. (1993). *Handbook of Individual Differences, Learning and Instruction*. Hillsdale, NJ: Erlbaum.
- Jones, D. M. (1948). Experiment in adaptation to individual differences. *Journal of Educational Psychology*, 39, 257-272.
- Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, 31, 268-289.
- Kalakoski, V. (2008). Effect of skill level on recall of visually presented patterns of musical patterns. *Scandinavian Journal of Psychology*, 48, 87-96.
- Kalyuga, S. (2006a). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology*, 20, 333-342.

- Kalyuga, S. (2006b). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology, 26*, 735-749.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539.
- Kalyuga, S. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnosing learner task-specific expertise. *Journal of Educational Psychology, 100*, 603-612.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5-23.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568.
- Kennedy, M. M. (1999). A test of some common contentions about educational research. *American Educational Research Journal, 36*, 511-541.
- Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. Arlington, VA: SRI international.
- Kulik, C. C., & Kulik, J. A. (1982). Meta-analysis of evaluation findings: Effects of ability grouping on secondary school students. *American Educational Research Journal, 19*, 415-428.
- Kulik, C. C. & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75-94.
- Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*, 73-77.

- Kulik, C.C., Kulik, J. A. & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60, 265-299.
- Lambiotte, J. G., & Dansereau, D. F. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *Journal of Experimental Education*, 60, 189-201.
- Lampert, M. (1985). How do teachers manage to teach? *Harvard Educational Review*, 55, 178-194.
- Langer, J. A., & Applebee, A. N. (1986). Reading and writing instruction: Toward a theory of teaching and learning. *Review of Research In Education*, 13, 171 – 194.
- Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology: Applied*, 11, 266-276.
- Leahy, W., & Sweller, J. (2007). The imagination effect increases with an increased intrinsic cognitive load. *Applied Cognitive Psychology*, 22, 273-283.
- Lepper, M. R. & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors in education* (pp. 135-138). New York: Academic Press.
- Levin, J. R., & O'Donnell, A. M. (1999). What to do about education research's credibility gaps? *Issues in Education*, 5, 177-192.
- 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, 423-458.
- 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, 101-111.
- Mandinach, E. B., & Corno, L. (1985). Cognitive engagement variations among students of different ability level and sex in a computer problem solving game. *Sex Roles*, 13, 241-251.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93, 187-198.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197-244.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1-43.

- Merriam-Webster, Inc. (2002). *Webster's new encyclopedia dictionary*. Springfield MA: Author.
- Merrill, D. C., Reiser, B. J., Merrill, S. K. & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction, 13*, 315-372.
- Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences, 2*, 277-305.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334.
- Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759.
- Nuthall, G. (2004). Relating classroom teaching to student learning: A critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review, 74*, 273-306.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education, 67*, 27-46.
- Palumbo, M. V., Miller, C. E., Shalin, V. J., & Steele-Johnson, D. (2005). The impact of job knowledge in the cognitive ability-performance relationship. *Applied H. R. M. Research, 10*, 13-20.
- Pascarella, E. T. (1978). Interactive effects of prior mathematics preparation and level of instructional support in college calculus. *American Educational Research Journal, 15*, 275-285.
- Pashler, H., McDaniel, M., Doug, R., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*, 105-119.
- Pellegrino, J. W., Baxter, G. P. & Glaser, R. (1999). Addressing the “two disciplines” problems: Linking theories of cognition and learning with assessment and instructional practice. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 307-353). Washington, DC: American Educational Research Association.

- Person, N. K., & Graesser, A. C. (2003) Fourteen facts about human tutoring: Food for thought for ITS developers. In *AI-ED 2003 workshop proceedings on tutorial dialogue systems: With a view towards the classroom* (pp. 335-344). Sydney, Australia: University of Sydney School of Information Technologies.
- Pressley, M., & Harris, K. R. (1990). What we really know about strategy instruction. *Educational Leadership, 48*, 31-34.
- Putnam, R. T. (1987). Structuring and adjusting content for tutees: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24*, 13-48.
- Quintilian. (1920). *Institutes of Oratory* (trans. H. E. Butler). London: William Heinemann.
- Randi, J., & Corno, L. (1997). Teachers as innovators. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *International handbook of teachers and teaching* (Vol.1, pp. 1163-1172). Dordrecht, the Netherlands: Kluwer Academic.
- Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.
- Regian, J. W., & Schneider, W. (1990). Assessment procedures for predicting and optimizing skill acquisition after extensive practice. In: N. Frederickson, R. Glaser, A. Lesgold, & M. G. Shafto (Eds), *Diagnostic Monitoring of Skill and Knowledge Acquisition* (pp. 297-323). Hillsdale, NJ: Erlbaum.
- Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education, 70*, 293-315.
- Ritter, G. W., Barnett, J. H., Denney, G. S. & Albin, G. R. (2009). The effectiveness of volunteer tutoring program for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79*, 3-38.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*, 836-852.
- Ross, S. M., & Rakow, E. A. (1981). Learner control versus program control as adaptive strategies for selection of instructional support on math rules. *Journal of Educational Psychology, 73*, 745-753.
- Salomon, G. (1974). Internalization of filmic schematic operations in interaction with learners' aptitudes. *Journal of Educational Psychology, 66*, 499-511.
- Sausner, R. (2005). Deceived, denied and in crisis. *District Administration, 41*, 26-31.

- Schaefer, P. S., Blankenbeckler, N. P., & Brogdon, C. J. (2011). *Measuring noncommissioned officer knowledge and experience to enable tailored training*. (Research Report 1952). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA552941)
- Schaefer, P. S., Blankenbeckler, N. P., & Lipinski, J. L. (2011). *Measuring officer knowledge and experience to enable tailored training*. (Research Report 1953). Arlington, VA. U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA552440)
- Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science, 1*, 89-92.
- Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2*, 8-9.
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.
- Shanahan, T. (1998). On the effectiveness and limitations of tutoring in reading. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of research in education* (Vol. 23, pp. 217-234). Washington, DC: American Educational Research Association.
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*, 159-189.
- Shen, C., & Tsai, H. (2009). Design principles of worked examples: A review of the empirical studies. *Journal of Instructional Psychology, 36*, 238-244.
- Shute, V. J. (1992). Aptitude-treatment interactions and cognitive skill diagnosis. In W. J. Regian & V. J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 15-43). Hillsdale, NJ: Erlbaum.
- Shute, V. J. (1993). A macroadaptive approach to tutoring. *Journal of Artificial Intelligence in Education, 4*, 61-93.
- Shute, V. J., Lajoie, S. P., & Gluck, K. A. (2000). Individualized and group approaches to training. In S. Tobias & J. D. Fletcher (Eds.), *Training and retraining: A handbook for business, industry, government and the military* (pp. 171-207). New York, NY: Macmillan.
- Shute, V. J., & Zapata-Rivera, D. (2008). Adaptive technologies. In J. M. Spector, D Merrill, J. van Merriënboer & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3<sup>rd</sup> ed., pp. 277-294). New York: Erlbaum.

- Siler, S. A. & Van Lehn, K. (2003). Accuracy of tutors' assessments of their students by tutoring context. In R. Alterman & D. Hirsch (Eds.), *Proceedings of the twenty-fifth annual meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, 50, 315-342.
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, 57, 293-336.
- Slavin, R. E. (1990). Ability grouping and student achievement in secondary schools: A best evidence synthesis. *Review of Educational Research*, 60, 471-490.
- Slavin, R. E. (1991). Synthesis of research on cooperative learning. *Educational Leadership*, February, 71-81.
- Slavin, R. E. (1993). Ability grouping in the middle grades: Achievement effects and alternatives. *The Elementary School Journal*, 93, 535-552.
- Smith, B., & Sechrest, L. (1991). Treatment of aptitude x treatment interactions. *Journal of Consulting and Clinical Psychology*, 59, 233-244.
- Snow R. E. (1976). Learning and individual differences. In Shulman, L. S. (Ed), *Review of research in education* (Vol. 4, pp. 50-105). Itasca, IL: F. E. Peacock.
- Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, 59, 205-216.
- Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, 27, 5-32.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.
- Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research*, 20, 241-254.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257-287.
- Tobias, S. (1989). Another look at research on the adaptation of instruction to student characteristics. *Educational Psychologist*, 24, 213-227.
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91, 334-341.

- Turney, A. H. (1931). The status of ability grouping. *Educational Administration and Supervision, 17*, 21-42, 110-127.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist, 46*, 197-221.
- Vineberg, R., & Taylor, E. N. (1972). *Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance*. Alexandria, VA: Human Resources Research Organization (DTIC No. ADA750603).
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wampler, R. L., Bink, M. L., & Cage, E. A. (2011). *Training aids for basic combat skills: Developing map-reading skills*. (Research Report 1941). Arlington, VA. U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA556404).
- Wang, M. C., & Walberg, H. J. (Eds.) *Adapting instruction to individual differences*. Berkeley, CA: McCutchan.
- Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74*, 642-655.
- Webb, N. M. (1984). Stability of small group interaction and achievement over time. *Journal of Educational Psychology, 76*, 211-224.
- Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.
- Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issue in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*, 607-651.
- Weinstein, R. S. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. *Journal of Educational Psychology, 88*, 103-116.
- Whitburn, J. (2001). Effective classroom organization in primary schools: Mathematics. *Oxford Review of Education, 27*, 411-428.
- Wickens, D. D. (1987). The dual meanings of context: Implications for research, theory, and applications. In D. S. Gorfein and R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference*. Hillsdale, NJ: Erlbaum.
- William, D., & Bartholomew, H. (2004). It's not which school but which set you're in that matters: The influence of ability grouping practices on student progress in mathematics. *British Educational Research Journal, 30*, 279-293.

- Wittwer, J., Nuckles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations? *Journal of Educational Psychology, 102*, 74-89.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.
- Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *Journal of Experimental Education, 67*, 197-217.
- Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*, 1-21.