ned, deliberate, and
:per understanding.
.e curricula, lead to
ıf some control over
: surface knowledge
leep understanding
ıost often there is a
one can *think about*
ıch emphasis on the
ge programs) is not
ɔn task, exposure to
portunities to prac-
·rograms need to be
ıe first time around.
: achieving students
:e is a need to iden-
ıd science), and this
dentifying what the
he child has or does

; most likely relates
o-learning, working
social skills training
n solving skills, self-
portant outcomes of
r that they could not
ı on the progression
of schooling. While
ıtexts, it is clear that
ıdventure programs).
ıoperation needed to
:rategies (particularly
eralized and used in
learning intentions,
:edback.

Chapter 9

The contributions from teaching approaches—part I

To keep the discussion on the various teaching approaches to a reasonable size, the contributions are divided, somewhat arbitrarily, into two chapters. The first chapter looks at goals, success criteria, and fostering student involvement, and the second other teaching approaches such as direct instruction, school-wide programs, using technology, and out-of school learning. This first of these two chapters follows a model of teaching and learning based on Clarke (2001; Clarke, Timperley, & Hattie, 2003), where the learning intentions and success criteria frame the challenge and purpose of the lesson. If such goal-directed lessons are to be successful, they must also use appropriate feedback, take account of students' views of the process of learning, and ensure students are actively involved in monitoring their own learning and developing their own meta-cognitive skills.

In a portrait of an exemplary school serving students who had been struggling to achieve and not enjoying schooling, Pressley, Gaskins, Solic, and Collins (2006) showed the power of teaching various learning strategies to these students. They claimed that when teachers critically reflected on conceptions of competent thinking and then taught various learning strategies to students, this was more likely to lead to engaging students in acquiring procedural and declarative knowledge and then to the students actually using this knowledge. The exemplary school emphasized the *engagement* of students in the learning process, teachers *articulating strategies of instruction* and paying attention to learning theories, and the school building as an infrastructure to support such instruction. The teachers provided constant scaffolding and modeling, attended to the day-to-day monitoring of students, and *sought feedback about their teaching* while also being concerned with making decisions about *optimal challenging tasks* to assign, and seeking insights from other professionals (e.g., counselors and mentors) about engaging students. There is much more, but the key ingredients of what it means to be strategic in teaching and learning relates to teachers finding ways to engage and motivate students, teach appropriate strategies in the context of various curricula domains, and constantly seeking feedback about how effective their teaching is being with all the students. The portrait by Pressley *et al.* sets the scene for this and the next chapter, which emphasizes the importance of setting challenging tasks, knowing when one (the teacher and the student) is successful in attaining these goals, the power of feedback, and the critical role of teaching appropriate learning strategies.

Table 9.1 Summary information from the meta-analyses on the contributions from teaching approaches

	No. metas	No. studies	No. people	No. effects	d	SE	CLE	Rank
Strategies emphasizing learning intentions								
Goals	11	604	41,342	820	0.56	0.057	40%	34
Behavioral organizers/advance organizers	11	577	3,905	1,933	0.41	0.040	29%	61
Concept mapping	6	287	8,471	332	0.57	0.051	40%	33
Learning hierarchies	1	24	—	24	0.19	—	13%	110
Strategies emphasizing success criteria								
Mastery learning	9	377	9,323	296	0.58	0.055	41%	29
Keller's PIS	3	263	—	162	0.53	—	37%	40
Worked examples	1	62	3,324	151	0.57	0.042	40%	30
Strategies emphasizing feedback								
Feedback	23	1,287	67,931	2,050	0.73	0.061	52%	10
Frequency or effects of testing	8	569	135,925	1,749	0.34	0.044	24%	79
Teaching test taking and coaching	10	267	15,772	364	0.22	0.024	16%	103
Providing formative evaluation	2	30	3,835	78	0.90	0.079	64%	3
Questioning	7	211	—	271	0.46	0.068	32%	53
Teacher immediacy	1	16	5,437	16	0.16	—	8%	115
Strategies emphasizing student perspectives in learning								
Time on task	4	100	—	136	0.38	0.101	27%	70
Spaced vs. massed practice	2	63	—	112	0.71	—	—	12
Peer tutoring	14	767	2,676	1,200	0.55	0.103	39%	36
Mentoring	2	74	10,250	74	0.15	0.047	11%	120
Strategies emphasizing student meta-cognitive/self-regulated learning								
Meta-cognitive strategies	2	63	5,028	143	0.69	0.181	49%	13
Study skills	14	668	29,311	2,217	0.59	0.090	41%	25
Self-verbalization/self-questioning	3	113	3,098	1,150	0.64	0.060	45%	18
Student control over learning	2	65	—	38	0.04	0.176	5%	132
Aptitude-treatment interactions	2	61	1,434	340	0.19	0.070	14%	108
Matching style of learning	8	411	29,911	1,218	0.41	0.016	29%	62
Individualized instruction	9	600	9,380	1,146	0.23	0.056	16%	100
Total	155	7,559	386,353	16,020	0.45	0.071	31%	—

Strategies emphasizing learning intentions

This section on learning intentions covers the teaching strategies of:

1 goals;
2 behavioral objectives;
3 organizers and adjunct questions;
4 concept mapping;
5 learning hierarchies.

Learning intentions describe what it is we want students to learn in terms of the skills, knowledge, attitudes, and values within any particular unit or lesson. Learning intentions should be clear, and provide guidance to the teacher about what to teach, help learners be

aware of what they should learn from the lesson, and form the basis for assessing what the students have learnt and for assessing what the teachers have taught well to each student. The activities planned for the lesson need to be focused on these intentions and move away from the all-too-often "busy" work that students might enjoy but which has little relationship to the learning intention.

Clarke, Timperley and Hattie (2003) have noted some important points about learning intentions and planning.

- Not all students in the class will be working at the same level, so it is important to adapt the learning intentions to make them appropriate to all students.
- The amount of time allocated should not be the same for all learning intentions, but should vary depending on whether they are developing concepts, skills or knowledge—*concepts or deeper learning* are likely to need more time than, say, the acquisition of *knowledge* or *surface information*.
- Learning intentions and activities can be grouped, because one activity can contribute to more than one learning intention, or one learning intention may need several activities or several exposures to the activities for the students to understand it fully.
- While learning intentions are what we intend students to learn, the students may also learn other things not planned for, and we need to be aware of these unintended consequences.

A more specific type of learning intention is the "mastery goal". Ames (1992) explained that, with a mastery goal, individuals are oriented toward developing new skills, trying to understand their work, improving their level of competence, or achieving a sense of mastery based on self-referenced standards. Elliott and Dweck (1988) further distinguished between mastery and learning goals. They defined learning goals as about more than the mastery of new things, and claimed that students encouraged to use learning goals were less worried about their intellect, remained focused on-task, and maintained their effective problem-solving strategies. Compatible with this goal construct is Brophy's (1983) description of "motivation to learn" whereby individuals focus on mastering and understanding content and demonstrate a willingness to engage in the process of learning.

Another important aspect of learning intentions is knowing *how* they will be implemented. Learning intentions take the form "I intend to reach x" and by articulating *how* they intend to reach "x", teachers and students are expressing an "implementation intention". Gollwitzer and Sheeran (2006) completed a meta-analysis testing the notion that implementation intentions help teachers and students attain goals. "Implementation intentions should enhance people's ability to initiate, maintain, disengage from, and undertake further goal pursuit and thereby increase the likelihood that strong goal intentions are realized successfully" (p. 20). They used 63 studies and the effect size was $d = 0.65$. It is not just the presence of a learning intention and having commitment that helps, but most importantly it is having a sense of "if-then" that helps the implementation of goal intentions. Thus, the art is setting appropriately challenging goals, developing commitment to attaining them, and developing intentions to implement strategies to attain them.

Goals

Locke and Latham (1990) have provided a compelling set of evidence, including many meta-analyses (but few with school achievement as the outcome) that indicate how critical

goals are for enhancing performance. They argued that goals serve a variety of functions that are essential in the teaching process: goals regulate action and they explain the nature of the link between the past and the future; and goals assume that human action is directed by conscious goals and intentions, although they do not assume that all human action is under fully conscious control (as we shall see later). A major finding of their book is that achievement is enhanced to the degree that students and teachers set challenging rather than "do your best" goals, relative to the students' present competencies.

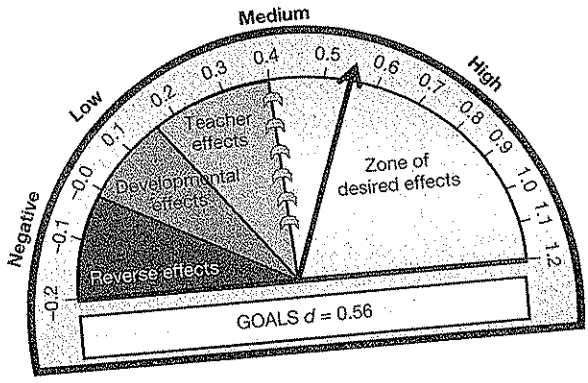
A major reason difficult goals are more effective is that they lead to a clearer notion of success and direct the student's attention to relevant behaviors or outcomes, whereas "doing your best" can fit with a very wide range of goals. It is not the specificity of the goals but the difficulty that is crucial to success. There is a direct linear relationship between the degree of goal difficulty and performance. There are five meta-analyses relative to this contention (Table 9.2) and the overall effect size is a large $d = 0.67$ (these are not all achievement outcomes and so are not included in the Appendices of this book). The performances of the students who have the most challenging goals are over 250 percent higher than the performances of the subjects with the easiest goals (Wood & Locke, 1997).

Also, difficult goals are much better than "do your best" or no assigned goals. Any school with the motto "do your best" should immediately change it to "face your challenges" or "strive to the highest". The following five meta-analyses relate to this contention (Table 9.3). This is because "do your best" goals are easily attained—in one sense, anything you do can be defined as your best. Instead, teachers and learners should be setting challenging goals.

Goals have a self-energizing effect if they are appropriately challenging for the student, as they can motivate students to exert effort in line with the difficulty or demands of the goal. Commitment to the goals helps, but is not necessary for goal attainment—except for special education students, where commitment makes a major difference. Klein, Wesson, Hollenbeck, and Alge (1999) found a high relationship ($d = 0.47$) between goal commitment and subsequent performance, and the effect between commitment and outcome increased as a function of goal difficulty. Donovan and Radosevich (1998) found lower effects of commitment to goals than they expected, but these were still quite high ($d = 0.36$).

Thus, goals inform individuals:

> as to what type or level of performance is to be attained so that they can direct and evaluate their actions and efforts accordingly. Feedback allows them to set reasonable



KEY	
Standard error	0.057 (Medium)
Rank	34th
Number of meta-analyses	11
Number of studies	604
Number of effects	820
Number of people (7)	41,342

Table 9.2 Relation between goal difficulty and performance

Authors	*Year*	*No. studies*	*No. effects*	*d*
Chidester & Grisgby	1984	12	1,770	0.52
Mento, Steel, & Karren	1987	70	7,407	0.55
Tubbs	1986	56	4,732	0.82
Wofford, Goodwin, & Premack	1992	3	207	0.90
Wood, Mento, & Locke	1987	72	7,548	0.58
Total	—	213	21,664	0.67

> goals and to track their performance in relation to their goals so that adjustments in effort, direction, and even strategy can be made as needed.
>
> (Locke & Latham, 1990, p. 23)

The scenario is that effective teachers set appropriately challenging goals and then structure situations so that students can reach these goals. If teachers can encourage students to share commitment to these challenging goals, and if they provide feedback on how to be successful in learning as one is working to achieve the goals, then goals are more likely to be attained.

Because assigned goals provide an individual with normative information on the expected level of performance, such goals have major effects on the development of self-efficacy and confidence, which in turn affects the choice of difficulty of goals. Table 9.4 provides a summary of meta-analyses as to the relationship between higher levels of self-efficacy and goal attainment (average $d = 0.92$).

A basis of many claims about the value of student self-assessment, self-evaluation, self-monitoring, and self-learning is that students have a reasonable understanding of where they are at, where they are going, what it will look like when they get there, and where they will go to next: that is, they have clear goals, learning intentions, and success criteria. Martin (2006) argued that one method to assist students to set task-specific and situation-specific goals was to use the notion of "personal bests". Task-specific goals provide students with clear information about what they are trying to achieve in the immediate future (both in terms of specificity and degree of challenge), and situation-specific goals provide students with the reason they want to achieve a particular outcome (to beat one's previous level of achievement on that goal). He found that setting personal bests had high positive relationships to educational aspirations, enjoyment of school, participation in

Table 9.3 Difficulty compared to "do your best" goals

Authors	*Year*	*No. studies*	*No. students*	*d*
Chidester & Grigsby	1984	17	2400	0.51
Guzzo, Jette, & Katzell	1985	na	na	0.65
Hunter & Schmidt	1983	17	1278	0.80
Mento, Steel, & Karren	1987	49	5844	0.42
Tubbs	1986	48	4960	0.50
Wood, Mento, & Locke	1987	53	6635	0.43
Total	—	184	21117	0.66

Table 9.4 Relation of self-efficacy to goal attainment

Study	Year	No. studies	*d*
Ajzen & Madden	1986	169	0.57
Ajzen & Madden	1986	90	0.44
Bandura & Cervone	1986	88	0.43
Garland	1985	127	0.39
Hollenbeck & Brief	1987	47	0.49
Locke, Frederick, Lee, & Bobko	1984	181	0.54
Meyer	1988	90	0.69
Meyer & Gellatly	1988	56	0.62
Meyer & Gellatly	1988	60	0.48
Silver & Greenhalis	1983	56	0.29
Taylor	1984	223	0.20
Weiss & Rakestraw	1988	80	0.60
Wofford, Goodwin, & Premack	1992	6	1.06
Wood & Locke	1987	517	0.32
Total	—	1784	0.46

class, and persistence on the task. The most salient features of the personal bests were the specificity and degree of challenge of the goals, and that the goals were seen to relate to self-improvement. Personal bests combined the best features of mastery and performance goals, as personal bests "primarily reflect a mastery orientation because it is self-referenced and self-improvement based and yet holds a slice of performance orientation because the student competes with his or her own previous performance" (Martin, 2006, p. 816).

Challenging goals are also effective when teaching special education students. Fuchs and Fuchs (1986) reported an effect of $d = 0.63$ for long-term and $d = 0.67$ for short-term goals. More importantly, there was an interaction effect with the outcome measure. For more probe-like outcomes the effect of challenge was largest for short-term goals ($d = 0.85$ compared to $d = 0.41$), whereas the reverse was the case for global outcomes ($d = 0.45$ for short-term and $d = 0.92$ for long-term goals). This indicates a need, therefore, to set appropriately challenging short-term goals for surface learning outcomes and set appropriately challenging long-term goals for deep learning outcomes.

It has been noted that "challenge" is a relative term—relative to a student's current performance and understanding, and to the success criteria deriving from the learning intention. The challenge should not be so difficult that the goal is seen as unattainable, given the student's level of self-efficacy or confidence; rather, teachers and students must be able to see a pathway to attaining the challenging goal—a pathway which can include strategies for understanding the goal or intention, implementation plans to attain it, and, preferably, a commitment to attaining the goal. Burns (2002) was specific: He used meta-analysis to ascertain the optimal ratio of known to unknown tasks for drill tasks (which is but one specific set of tasks that teachers can engage students with). He found that the ratios differed depending on whether the student was in the acquisition or proficiency stage (the former relates to acquiring the knowledge and information, the latter relates to increasing accuracy and fluency). He also acknowledged that there was a maintenance, generalization, or application stage but there were no studies investigating the appropriate ratios at this stage. Drill ratios were more applicable to the acquisition ($d = 1.09$), than to the proficiency ($d = 0.39$) stage; and the optimal rate seems to be to include at least 90

percent known to unknown items in the tasks (d = 1.19) and certainly not less than 50 percent known to unknown (d = 0.49). Gickling (1984) showed that the ratios for learning to read needed to be more like ninety-five percent known to five percent unknown words in a text. It is also important for the teacher to choose the tasks with these ratios, as the effects are much greater than when students choose the ratios. While not explored, there are suggestions that the ratios may need to be higher when deeper learning is desired rather than surface knowledge.

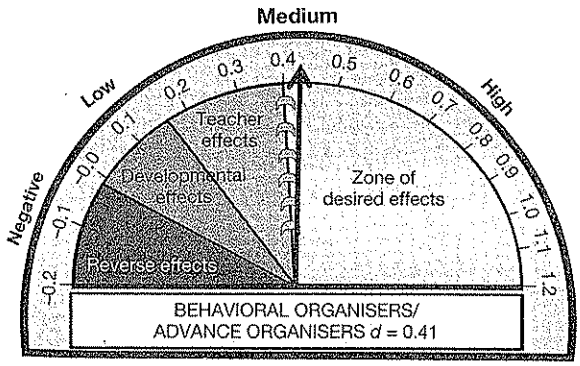
Behavioral objectives and advance organizers

Advance organizers can be:

> broadly defined as bridges from the reader's previous knowledge to whatever is to be learned; they are supposed to be more abstract and inclusive than the more specific material to be learned, and to provide a means for organizing the new material.
>
> (Stone, 1983, p. 194)

They are aimed to bridge and link old with new information, and as they are meant to be presented prior to learning, then advance organizers can assist in helping the learner organize and interpret new upcoming instruction. Similarly, behavioral objectives are statements of what students ought to be able to do as a consequence of instruction (Popham, Eisner, Sullivan, & Tyler, 1969), but they tend to be more often used for surface rather than deeper knowledge. The overall effects show much variance but the effects are highest when the learning intentions of the lessons are articulated, when notions of success included, and when these are shared with the students. When they are primarily for the teacher, usually in lesson plans, or aimed primarily at surface learning and not including any deep learning, then the effects are lower. Kozlow (1978) found that behavioral objectives were more effective when they involved comparisons to some standards of performance rather than being expository in nature.

Luiten, Ames, and Ackerman (1980) found that advance organizers have a small but facilitative effect on both learning and retention, with the effect increasing over time (d = 0.21). Similarly, Stone (1983) found that advance organizers were associated with increased learning and retention of teaching material. Using advance organizers to introduce new material, by providing a bridge from previous knowledge, did facilitate long-term learning,



KEY	
Standard error	0.040 (Low)
Rank	61st
Number of meta-analyses	11
Number of studies	577
Number of effects	1,933
Number of people (2)	3,905

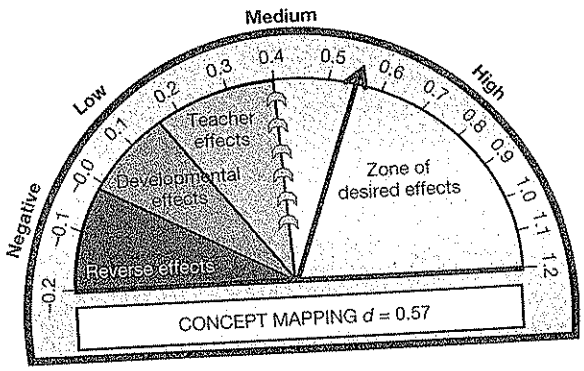
but the effects were lower for written advance organizers compared to non-written ones, and had no effect when used for teaching low-ability or low-knowledge learners. Too often, advance organizers and behavioral objectives tended to be specific, ignore challenge, and have no notions of what would be deemed as success in attaining the objective.

Concept mapping

Concept mapping involves the development of graphical representations of the conceptual structure of the content to be learnt. Thus, it can be considered as a form of learning intention, if for no other reason than it identifies the material to be learnt, oftentimes with indicators of priorities and higher-order concepts. As with behavioral objectives and learning hierarchies, concept mapping derives from Ausubel's (1968) claims that concepts can be organized in hierarchical form in the cognitive structure, and it helps learning if concepts related to what is to be learned can be linked to the concept maps a student already has (see also Novak, 1977). The difference between concept mapping and other organizing methods (e.g., behavioral objectives, learning hierarchies) is that it involves the students in the development of the organizational tool.

The importance of concept mapping relates to its emphasis on summarizing the main ideas in what is to be learnt—although only if the students have some familiarity with the surface knowledge of the (often deeper) concept to be mapped. Concept mapping can assist in synthesizing and identifying the major ideas, themes, and interrelationships—particularly for the learners who do not have these organizing and synthesizing skills. Kim, Vaughn, Wanzek, and Wei (2004) argued that the visual displays of information such as those provided by concept mapping enhance the reading comprehension of students with learning difficulties, possibly by helping these students organize the verbal information and thereby improving their recall.

Moore and Readance (1984) reported greater effects when concept mapping occurred after initial exposure to the material to be mapped (and not before or during this learning; see also Kang, 2002). Nesbit and Adesope (2006) found greater effects when the emphasis was on understanding the central rather than the detailed ideas of the topic being mapped. Nesbit and Adesope also found that there was little difference between concept mapping and asking students to construct an outline of the topic ($d = 0.19$), but the effects were larger for concept mapping when compared to lectures or discussions on the topic ($d = 0.74$). It is the heuristic process of organizing and synthesizing that is the important feature, and



KEY	
Standard error	0.051 (Medium)
Rank	33rd
Number of meta-analyses	6
Number of studies	287
Number of effects	332
Number of people (3)	8,471

concept mapping is but one of many of these methods—but an effective method. It does not seem to matter who does the mapping (student alone, in groups, or teacher, Horton *et al.*, 1993) but the strongest effects are when students provided the terms for the maps, regardless of who then devised the maps. Kim *et al.* (2004), however, found higher effects for teacher- than student-generated maps, whereas Nesbit and Adesope (2006) found higher effects when students were made to construct (d = 0.81), rather than just study, concept maps (d = 0.37).

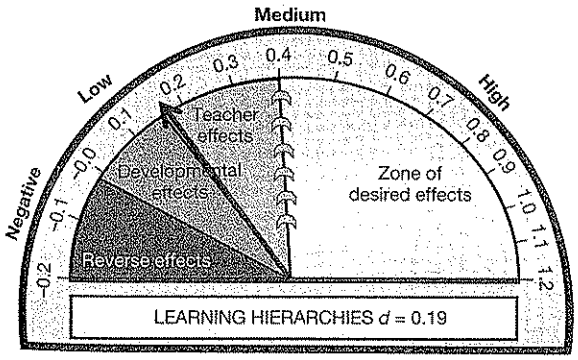
Various authors have found that the effects were highest with those students least likely to know the relationship between lower and higher-order notions; that is, with lower rather than higher ability or highly verbal students (Horton *et al.*, 1993; Nesbit & Adesope, 2006; Vásquez & Caraballo, 1993). As Nesbit and Adesope (2006) concluded, many of these gains may be "due to greater learner engagement occasioned by concept mapping ... rather than the properties of the concept map as an information medium" (p. 434), although it is noted that the effects from concept mapping were higher than for studying text passages, lists, and outlines. Thus they argue that it is not just the "summarizing/integrating" nature of concept maps, but also there may be a lower cognitive load "by arranging nodes in two-dimensional space to represent relatedness, consolidating all references to a concept in a single symbol, and explicitly labeling links to identify relationships" (p. 434).

Learning hierarchies

A different form of learning intention is to structure the learning in some form of hierarchy, such that it is more effective to acquire first a series of skills that will support later learning. Horon and Lynn (1980) found that learning hierarchies can facilitate learning (d = 0.19) and shorten learning time to a small extent (d = 0.09). Hierarchical instruction is more effective in promoting learning at the elementary level (d = 0.44) than at the high school level (d = 0.07). The overall effects are very low.

Strategies emphasizing success criteria

The purpose of the success criteria, or "What are we looking for?" is to make students understand what the teacher is using as the criteria for judging their work, and, of course, to ensure that the teacher is clear about the criteria that will determine if the learning intentions have been successfully achieved. Too often students may know the learning

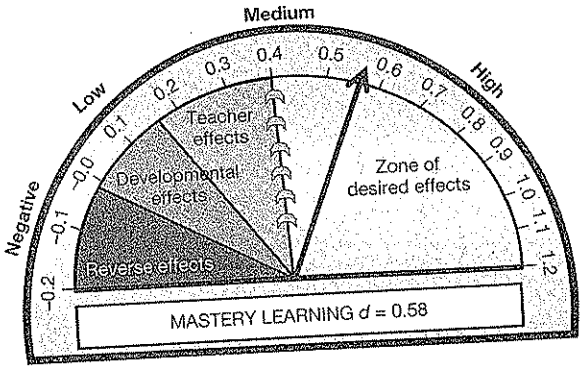


KEY	
Standard error	na
Rank	110th
Number of meta-analyses	1
Number of studies	24
Number of effects	24
Number of people (0)	na

intention, but not how the teacher is going to judge their performance, or when or whether they have been successful. A learning intention of "to learn to use effective adjectives", for instance, does not give the students the marking criteria or how they will be judged. The success criteria, or "How will we know?", need to state as exactly as possible what the students and teacher will want to see. In this case, two alternatives might be: "What you're looking for is that you have used at least five effective adjectives", or "What you're looking for is that you have used an adjective just before a noun on at least four occasions that will help to paint a detailed picture so the reader can understand the feel of the jungle and the light of the jungle". It is important that the success criteria are as clear and specific as possible (at surface or at deep levels, or both) because the teacher (and learner) needs to monitor the students' progress throughout the lesson to make sure they understand the intended meaning. There are three sets of related notions that emphasize success criteria: mastery learning, Keller's personalized system of instruction, and the provision of worked examples.

Mastery learning

The claim underlying mastery learning is that all children can learn when provided with clear explanations of what it means to "master" the material being taught. Other features involved include: appropriate learning conditions in the classroom, such as high levels of cooperation between classmates; high levels of teacher feedback that is both frequent and specific by using diagnostic formative tests; and the regular correction of mistakes students make as they travel along their learning path. Mastery learning requires numerous feedback loops, based on small units of well-defined, appropriately sequenced outcomes. Bloom (1968) defined mastery in terms of behavioral objectives, with class instruction supplemented by feedback or correction mechanisms. Willett, Yamashita, and Anderson (1983, p. 408) added that "tests on unit objectives are followed by supplementary instruction on objectives not attained, and the specific levels of attainment are specified". The important variable in mastery learning is the time required to reach the levels of attainment. The notion is that learning should be held constant and time should be allowed to vary, rather than the opposite, which is the norm in traditional instruction. The teacher determines the pace of the instruction and directs the accompanying feedback and corrective procedures (Guskey & Pigott, 1988). The material is divided into relatively small learning units, each with their own objectives and assessment. Each unit is preceded by brief diagnostic tests,



KEY	
Standard error	0.055 (Medium)
Rank	29th
Number of meta-analyses	9
Number of studies	377
Number of effects	296
Number of people (2)	9,323

which provide information to identify gaps and strengths. No student proceeds to new material until prior or more basic prerequisite material is mastered.

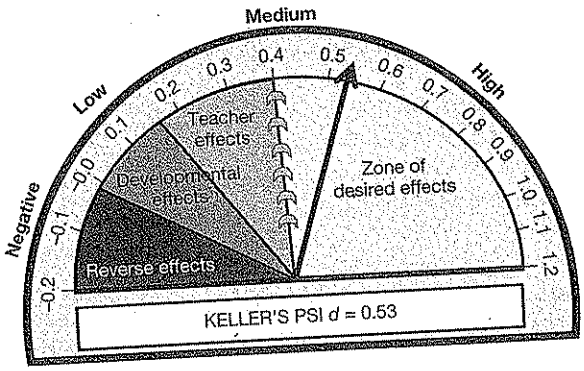
Willett *et al.* (1983) reviewed a dozen different innovations in teaching strategies, and mastery learning had the highest effects. They argued that mastery learning was the most successful innovative system, closely followed my Keller's PSI (see next section). Guskey and Gates (1986) found similar high effects for mastery learning in each of elementary school ($d = 0.94$), high school ($d = 0.72$), and college ($d = 0.65$). In a follow-up study, Guskey and Piggott (1988), using group-based applications of mastery strategies, showed consistently positive effects on both cognitive and affective student learning outcomes. Kulik and Kulik (1986) determined that testing for mastery had positive effects on student achievement both at college and pre-college levels ($d = 0.52$). The effects of mastery testing were particularly strong on lower ability students ($d = 0.96$). Mastery testing, they argued, increased the amount of instructional time required by, on average, 25 percent. Their evidence, however, did not support Bloom's prediction that variation in performance will be reduced to near zero with mastery testing procedures.

Kulik, Kulik, and Bangert-Drowns (1990) found mastery learning programs had a positive effect on examination performance of students in colleges, high schools, and the upper grades of elementary schools, raising examination performance by about half a standard deviation, especially for low-aptitude students. Mastery programs had positive effects on student attitudes towards course content and instruction, but increased student time spent on instructional tasks. Self-paced mastery programs often reduced completion rates in college classes.

The only exception to the positive findings on mastery learning programs is the meta-analysis by Slavin (1987), who found no evidence to support the effectiveness of group-based mastery learning on standardized achievement measures. One of the features of Slavin's argument is that studies that do not meet his criteria should be excluded, which leaves only seven articles—a very small representation of a large set of potential studies. His criteria included: students had to have been tested on their mastery at least once every four weeks, only studies where students were taught as a total group were included, studies could not use a feedback-corrective cycle, interventions had to last a minimum of four weeks, and there had to be at least two experimental and two control groups used.

Keller's Personalized System of Instruction

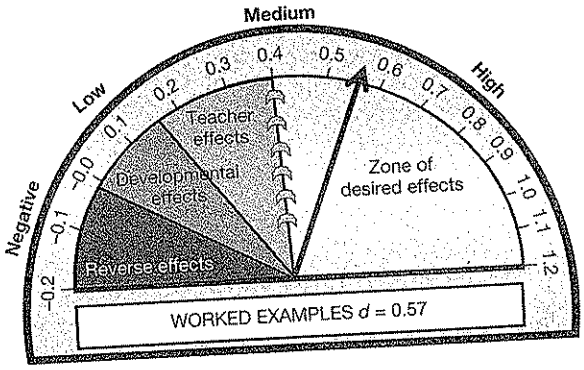
A specific implementation of mastery learning is the Personalized System of Instruction, developed by Keller and Sherman during the 1960s as a form of programmed instruction that employs a highly structured, student-centered approach to course design that emphasizes self-pacing and mastery (Keller, 1968; Keller & Sherman, 1974). The key features include: students proceed through the course at their own pace; students demonstrate mastery of each component of the course before proceeding to the next; teaching materials and other communications between teachers and students are largely text-based; and teachers are involved more in tutorial support and in providing motivation for students to complete the work and attain the goals. The effects are very similar to the other mastery learning programs. The meta-analyses show that students using PSI had higher grades and higher satisfaction rates than students in conventional classes, but that study time was similar in both types of classes (Kulik, Kulik, & Cohen, 1980).



KEY	
Standard error	Na
Rank	40th
Number of meta-analyses	3
Number of studies	263
Number of effects	162
Number of people (0)	na

Worked examples

Another form for demonstrating to students what "success" looks like and thus what the goal could be for their own learning, is by providing them with worked examples (Crissman, 2006). Worked examples typically consist of a problem statement and the appropriate steps to the solution. The defense for providing such worked examples is that they reduce the cognitive load for students such that they concentrate on the processes that lead to the correct answer and not just providing an answer (which may or may not be correct). A typical example of worked examples consists of three parts: an introductory phase (exposure to the example), an acquisition or training phase, and a test phase (assessing the learning). Most studies follow this pattern, although there may be slight deviations, such as the inclusion of a pretest or the introduction of a delayed acquisition or delayed test phase, or both. The studies used for this meta-analysis involved the use of worked examples to alleviate cognitive load in the learner. The overall effect was $d = 0.52$, and most programs were close to this average: intra-example features (such as multiple examples, story variation, example/problem pairs) had an effect size of $d = 0.52$; the effect size for conventional worked examples was $d = 0.49$; integration of sources of information (e.g., diagrams, text) was $d = 0.52$; fading (omitting some of the steps in the example) was $d = 0.60$; inclusion of subgoals was $d = 0.52$; and self-explanations of the steps as they used the worked example was $d = 0.57$. All these various types of instruction using worked examples generally help to reduce cognitive load.



KEY	
Standard error	0.042 (Medium)
Rank	30th
Number of meta-analyses	1
Number of studies	62
Number of effects	151
Number of people (1)	3,324

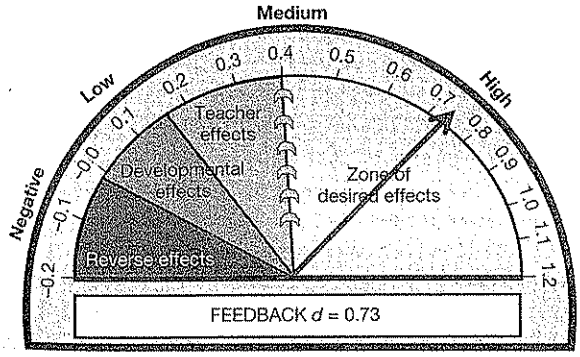
There do seem to be worthwhile effects from providing worked examples to students, but it is more difficult to find evidence of the effects from providing worked examples to teachers (often called exemplars). Peddie, Hattie and Vaughan (1999) completed an exhaustive search of evidence for research on the effects of exemplars and could find much rhetoric and many claims about their importance. When 50-plus organizations that had developed exemplars were asked to send their research, all sent boxes of exemplars, but none were able to send evidence of their effects.

Implementations that emphasize feedback

This section outlines the meanings of feedback, the effects of different types of feedback, feedback via frequent testing, teaching test-taking skills, providing formative evaluation to teachers, questioning to provide teachers and students with feedback, and the immediacy of feedback.

Feedback

When I completed the first synthesis of 134 meta-analyses of all possible influences on achievement (Hattie, 1992) it soon became clear that feedback was among the most powerful influences on achievement. Most programs and methods that worked best were based on heavy dollops of feedback. When I was presenting these early results in Hong Kong, a questioner asked what was meant by feedback, and I have struggled to understand the concept of feedback ever since. I have spent many hours in classrooms (noting its absence, despite the claims of the best of teachers that they are constantly engaged in providing feedback), worked with students to increase self-helping (with little success), and have tried different methods of providing feedback. The mistake I was making was seeing feedback as something *teachers provided to students*—they typically did not, although they made claims that they did it all the time, and most of the feedback they did provide was social and behavioral. It was only when I discovered that feedback was most powerful when it is from the *student to the teacher* that I started to understand it better. When teachers seek, or at least are open to, feedback from students as to what students know, what they understand, where they make errors, when they have misconceptions, when they are not engaged—then teaching and learning can be synchronized and powerful. Feedback to teachers helps make learning visible.



KEY	
Standard error	0.061 (Medium)
Rank	10th
Number of meta-analyses	23
Number of studies	1,287
Number of effects	2,050
Number of people (10)	67,931

Recently a colleague and I published a paper devoted to the power of feedback, which provides a deeper explanation than can be presented in this book (Hattie & Timperley, 2007). But, in summary, feedback is information provided by an agent (e.g., teacher, peer, book, parent, or one's own experience) about aspects of one's performance or understanding. For example, a teacher or parent can provide corrective information, a peer can provide an alternative strategy, a book can provide information to clarify ideas, a parent can provide encouragement, and a learner can look up the answer to evaluate the correctness of a response. *Feedback is a "consequence" of performance.*

To assist in understanding the purpose, effects, and types of feedback, it is useful to consider a continuum of instruction and feedback. At one end of the continuum is a clear distinction between providing instruction and providing feedback. However, when feedback is combined with a correctional review, feedback and instruction become intertwined until "the process itself takes on the forms of new instruction, rather than informing the student solely about correctness" (Kulhavy, 1977, p. 212). To take on this instructional purpose, feedback needs to provide information specifically relating to the task or process of learning that fills a gap between what is understood and what is aimed to be understood (Sadler, 1989), and it can do this in a number of different ways. For example, this may be through affective processes, such as increased effort, motivation, or engagement. Alternatively, the gap may be reduced through a number of different cognitive processes, including helping students to come to a different viewpoint, confirming to the student that they are correct or incorrect, indicating that more information is available or needed, pointing to directions that the student could pursue, and indicating alternative strategies to understand particular information. Winne and Butler (1994) provided an excellent summary in their claim that "feedback is information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies" (p. 5740).

The effect sizes reported in the feedback meta-analyses show considerable variability, which indicates that some types of feedback are more powerful than others. The most effective forms of feedback provide cues or reinforcement to the learner, are in the form of video, audio or computer-assisted instruction feedback, or relate feedback to learning goals. It is also worth noting that the key is feedback that is received and acted upon by students—many teachers claim they provide ample amounts of feedback but the issue is whether students receive and interpret the information in the feedback. At best, each student receives moments of feedback in a single day (Nuthall, 2005; Sirotnik, 1983). Carless (2006) asked students and teachers whether teachers provided detailed feedback that helped students improve their next assignments. About 70 percent of the teachers claimed they provided such detailed feedback often or always, but only 45 percent of students agreed with their teachers' claims. Further, Nuthall (2005) found that most feedback that students obtained in any day in classrooms was from other students, and most of this feedback was incorrect.

Programmed instruction, praise, punishment, and extrinsic rewards were the least effective forms of feedback for enhancing achievement. Indeed, it is doubtful whether rewards should be thought of as feedback at all. Deci, Koestner, and Ryan (1999) have described tangible rewards (stickers, awards, and so on) as contingencies to activities rather than feedback because they contain so little task information. In their meta-analysis of the effects of feedback on motivation, these authors found a negative correlation between extrinsic

rewards and task performance ($d = -0.34$). Tangible rewards significantly undermined intrinsic motivation, particularly for interesting tasks ($d = -0.68$) compared to uninteresting tasks ($d = 0.18$). In addition, when the feedback was administered in a controlling manner (e.g., saying that the student performed as they "should" have performed), the effects were even worse ($d = -0.78$). Thus, Deci *et al.* concluded that extrinsic rewards are typically negative because they "undermine people's taking responsibility for motivating or regulating themselves" (Deci *et al.*, 1999, p. 659). Rather, extrinsic rewards are a controlling strategy that often leads to greater surveillance, evaluation, and competition, all of which have been found to undermine enhanced engagement and regulation (Deci & Ryan, 1985).

Providing feedback is not about giving rewards, but rather providing information about the task. Cameron and Pierce (1994) asked about the causal effects of extrinsic rewards and reinforcement on intrinsic motivation (hence this meta-analysis is not included in the Appendices because achievement is not the outcome). The results show that rewards did not significantly affect intrinsic motivation: the effects of rewards were $d = -0.06$ for free time on task, $d = 0.21$ for attitude, $d = 0.08$ for performance during free-time period, and $d = 0.05$ for willingness to volunteer. When intrinsic motivation was measured by attitude toward a task, rewarded subjects reported higher intrinsic motivation than non-rewarded subjects. Verbal rewards appeared to produce a positive effect and tangible rewards suggested a negative effect. Those rewarded with verbal praise or positive feedback showed greater intrinsic motivation and spent more time on a task once the reward was withdrawn than non-rewarded subjects. It is critical, however, to note how small these effects are and thus to conclude that rewards and praise are or are not critical seems moot.

The most systematic study addressing the effects of various types of feedback was published by Kluger and DeNisi (1996). Their meta-analysis included studies of feedback interventions that were not confounded with other manipulations, included at least a control group, measured performance, and included at least ten participants. Although many of their studies were not classroom or achievement based, their messages are of much interest. From the 131 studies, they estimated 470 effect sizes, based on 12,652 participants, and the average effect size was $d = 0.38$, and 32 percent of the effects were negative. Specifically, feedback is more effective when it provides information on correct rather than incorrect responses and when it builds on changes from previous trails. The impact of feedback was also influenced by the difficulty of goals and tasks. There is highest impact when goals are specific and challenging but when task complexity is low. Giving praise for completing a task appears to be ineffective, which is hardly surprising because it contains such little learning-related information. Feedback is more effective when there are perceived low rather than high levels of threat to self-esteem, presumably because low threat conditions allow attention to be paid to the feedback.

Figure 9.9 presents a framework in which feedback can be considered. The claim is made that the main purpose of feedback is to reduce discrepancies between current understandings and performance and a learning intention or goal. The strategies that students and teachers use to reduce this discrepancy depend partly on the level at which the feedback operates. These levels include the level of task performance, the level of process of understanding how to do a task, the regulatory or meta-cognitive process level, and the self or person (unrelated to the specifics of the task). Feedback has differing effects across these levels.

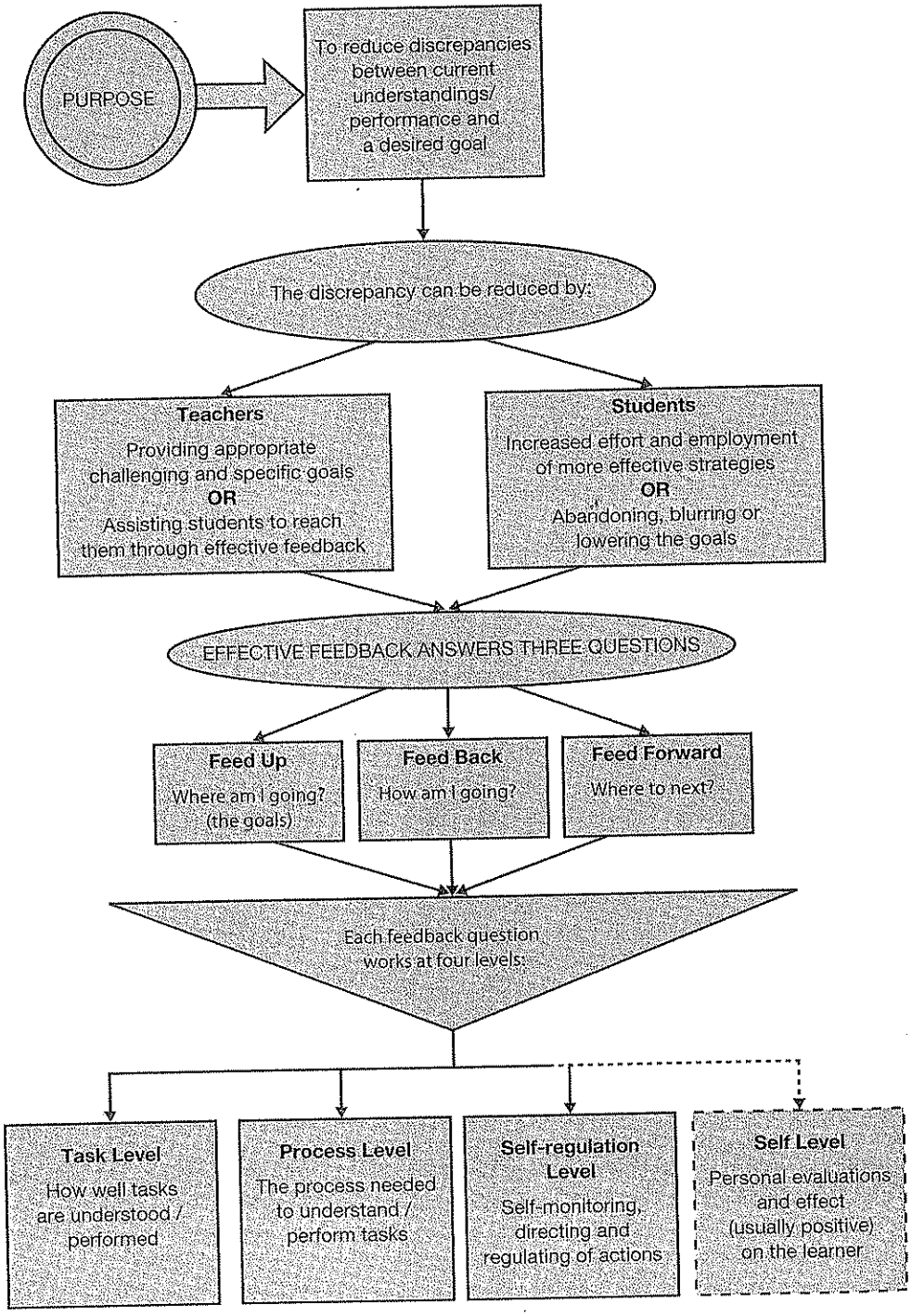


Figure 9.9 A model of feedback

The major :
success criteria
to next?" (prog
both teachers a
do not work in
relating to "Ho
next?" and "W
the gap betwee
power of feedb

So far so goo
four levels note
is correct or in
different, or co
of Versailles". S
or complete th
information, or
example, a teach
attending to the
of your meanin
strategies we tal
self-regulation I
further on the t
an argument. Cl
Such feedback c
self-beliefs abou
how to better an
in the sense tha
unrelated to perf
student", "Well c

The art is to p
student is worki
praise) is rarely e
tions and so is in
self, students try
minimize effort,
to minimize the
to the processes
beyond it to mo
and greater inves
fluency and mast

We need to be
teaching and lear
learners at the ac
elaborations throu
If feedback is dir
develop effective
tive, feedback nee

The major feedback questions are "Where am I going?" (learning intentions/goals/success criteria), "How am I going?" (self-assessment and self-evaluation), and "Where to next?" (progression, new goals). An ideal learning environment or experience is when both teachers and students seek answers to each of these questions. These three questions do not work in isolation at each of the four levels, but typically work together. Feedback relating to "How am I going?" has the power to lead to doing further tasks or "Where to next?" and "Where am I going?". As Sadler (1989) has convincingly argued, it is closing the gap between where the student is and where they are aiming to be that leads to the power of feedback.

So far so good, but the difficulty arises from the way in which feedback works at the four levels noted above. First, feedback can be about the task or product, such as the work is correct or incorrect. This level of feedback may include directions to acquire more, different, or correct information, such as "You need to include more about the Treaty of Versailles". Second, feedback can be aimed at the process used to create the product or complete the task. This kind of feedback is more directly aimed at the processing of information, or learning processes required for understanding or completing the task. For example, a teacher or peer may say to a learner, "You need to edit this piece of writing by attending to the descriptors you have used, so the reader is able to understand the nuances of your meaning", or "This page may make more sense if you use the comprehension strategies we talked about earlier". Third, feedback to the student can be focused at the self-regulation level, including greater skill in self-evaluation, or confidence to engage further on the task. For example, "You already know the key features of the opening of an argument. Check to see whether you have incorporated them in your first paragraph." Such feedback can have major influences on self-efficacy, self-regulatory proficiencies, and self-beliefs about the student as a learner, such that the student is encouraged or informed how to better and more effortlessly continue on the task. Fourth, feedback can be personal in the sense that it is directed to the "self" which, it will be argued below, is too often unrelated to performance on the task. Examples of such feedback include, "You are a great student", "Well done!".

The art is to provide the right form of feedback at, or just above, the level where the student is working—with one exception. Feedback at the self or personal level (usually praise) is rarely effective. Praise is rarely directed at addressing the three feedback questions and so is ineffective in enhancing learning. When feedback draws attention to the self, students try to avoid the risks involved in tackling a challenging assignment, they minimize effort, and they have a high fear of failure (Black & Wiliam, 1998) in order to minimize the risk to the self. Thus, ideally, teaching and learning move from the task to the processes and understandings necessary to learn the task, and then to continuing beyond it to more challenging tasks and goals. This process results in higher confidence and greater investment of effort. This flow typically occurs as the student gains greater fluency and mastery.

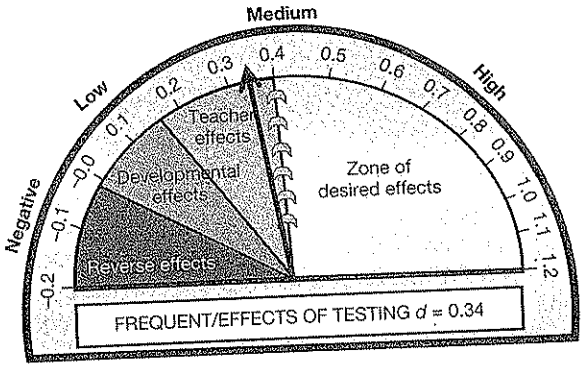
We need to be somewhat cautious, however. Feedback is not "the answer" to effective teaching and learning; rather it is but one powerful answer. With inefficient learners or learners at the acquisition (not proficiency) phase, it is better for a teacher to provide elaborations through instruction than to provide feedback on poorly understood concepts. If feedback is directed at the right level it can assist students to comprehend, engage, or develop effective strategies to process the information intended to the learnt. To be effective, feedback needs to be clear, purposeful, meaningful and compatible with students' prior

knowledge, and to provide logical connections. It also needs to prompt active information processing on the part of the learner, have low task complexity, relate to specific and clear goals, and provide little threat to the person at the self level. The major discriminator is whether feedback is clearly directed to the various levels of task, processes, or regulation, and not directed to the level of "self". These conditions highlight the importance of classroom climates that foster peer and self-assessment, and allow for learning from mistakes. We need classes that develop the courage to err.

Thus, when feedback is combined with effective instruction in classrooms, it can be very powerful in enhancing learning. As Kluger and DeNisi (1996) noted, a feedback intervention provided for a familiar task that contains cues that support learning, attracts attention to feedback–standard discrepancies at the task level, and is void of cues that direct attention to the self, is likely to yield impressive gains in students' performance. It is important to note, however, that under particular circumstances, instruction is more effective than feedback. Feedback can only build on something; it is of little use when there is no initial learning or surface information. In summary, feedback is what happens second, is one of the most powerful influences on learning, occurs too rarely, and needs to be more fully researched by qualitatively and quantitatively investigating how feedback works in the classroom and learning process.

Frequent testing/Effects of testing

Another form of feedback is repeated testing, but this is only effective if there is feedback from the tests to teachers such that they modify their instruction to attend to the strengths and gaps in student performance. Although performance is increased with more frequent testing, the amount of improvement in achievement diminishes as the number of tests increase (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Students taking at least one test during a 15-week term scored about half a standard deviation higher in criterion examinations than students taking no tests. When two groups answered identical test items, superior performance was obtained from students who answered the questions on a large number of short tests rather than on a small number of long tests. The caution is that it may not be the frequency of test taking but that frequent test taking made the learning intentions and success criteria more specific and transparent. Clariana and Koul (2006) found that multiple-try feedback was less effective for surface outcomes (d = –0.22) but more effective for higher-order outcomes (d = 0.10). "Multiple try



KEY	
Standard error	0.044 (Medium)
Rank	79th
Number of meta-analyses	8
Number of studies	569
Number of effects	1,749
Number of people (2)	135,925

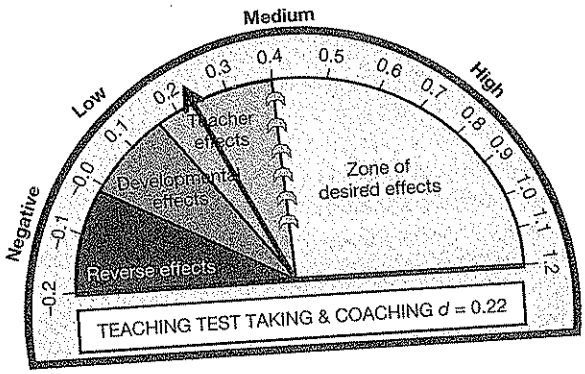
feedback on error requires the learner to think more about the lesson question, unless the learner just guesses randomly due to frustration or impatience" (p. 261). Similarly, Kim (2005) found that performance assessment was more effective the longer the period it had been implemented—as then students and teachers become more adept at completing this form of assessment.

The effect is not merely from testing and testing, it is from learning from testing. Gocmen (2003), for example, found an effect size of $d = 0.41$ from frequent testing, but this was higher when the testing was accompanied by feedback ($d = 0.62$) compared to no feedback ($d = 0.30$). Lee (2006) investigated the effects of statewide high-stakes testing and test-driven accountability policies on reading and mathematics achievement in the United States (since 1990). He found a $d = 0.36$ effect ($d = 0.29$ for reading and $d = 0.38$ for mathematics), but the effects only occurred in elementary ($d = 0.44$) and middle schools ($d = 0.35$) and not in high schools ($d = 0.03$). States with the strongest accountability programs made greater gains over the years than those with weaker accountability measures, but Lee noted that these gains mapped to similar trajectories from the years before these accountability policies were brought into law! He concluded that "to argue that states adopting strong accountability policies significantly improved student achievement is not convincing until substantial improvements in schooling conditions and practices occur" (p. 26).

Many states in the United States have high-stakes testing and there is also much testing embedded in the No Child Left Behind imperatives. There have been arguments that such frequent testing is akin to a coaching effect, whereas others consider that any gains are because of narrowing the curriculum, teaching to the test, and because too many students are excluded who may not perform so well. Amrein and Berliner (2002) raised much debate with their analysis of the performance of 18 states with high-stakes testing systems and found little effect of these systems on student achievement. This conclusion was contested (e.g., Braun, 2004; Raymond & Hanushek, 2003; Rosenshine, 2003). Lee (2006) used meta-analysis to compare different state policies on the National Assessment of Educational Progress examination. He found six studies favored high-stakes testing states, five were mixed, and one favored low-stakes testing states. The effects were extremely varied ($d = -0.67$ to $d = 1.24$), although it made no difference as to the focus of the accountability—that is, whether the focus is a combination of schools and students $d = 0.38$, for schools alone $d = 0.39$, or for students alone $d = 0.31$. The effects on mathematics ($d = 0.38$) are slightly higher than on reading ($d = 0.29$), and higher for elementary ($d = 0.44$) and middle schools ($d = 0.35$) than for high schools ($d = 0.03$).

Teaching test taking and coaching

The term "coaching" is used to refer to a wide range of test preparation activities carried out in order to improve test scores. Typically, coaching is instruction given or practice undertaken in preparation for taking a test (Cole, 1982). DerSimonian and Laird (1983) evaluated the effect of coaching on Scholastic Aptitude test scores and found that while the results did support the positive effect of coaching on SAT scores, the size of the coaching effect from the matched or randomized studies appeared too small to be practically important. Uncontrolled studies showed more variation in the effects attributed to coaching than matched or randomized studies and higher levels overall.



KEY	
Standard error	0.024 (Low)
Rank	103rd
Number of meta-analyses	10
Number of studies	267
Number of effects	364
Number of people (1)	15,772

Bangert-Drowns, Kulik, and Kulik, (1983) found the effects of coaching raised achievement test scores by $d = 0.25$. The level of intervention influenced effect sizes, with effect sizes smaller for short test-taking sessions, larger for more extensive programs, and greatest in single length programs designed to influence broader cognitive skills. An examination of 14 studies on the effectiveness of coaching for aptitude tests (Kulik, Bangert-Drowns, & Kulik, 1984) found that coaching raised scores on SAT as well as intelligence and other aptitude tests. SAT scores were raised $d = 0.15$ standard deviations with scores for aptitude and intelligence tests raised $d = 0.43$ standard deviation. The length of the training program also seems important. Samson (1985) reported that programs continuing for five weeks or more produced greater effects than those of a shorter duration. Samson also noted that the effects were higher with students in upper grades rather than in lower grade levels, and for students from lower socioeconomic backgrounds.

Hausknecht, Halpert, Di Paolo, and Gerrard, 2007 found an overall effect of $d = 0.22$ when test were re-administered, but much less for a third administration of the test. More specifically they found that the magnitude of practice was positively related to the amount of student contact time with coaching ($d = 0.26$), was greater for identical test ($d = 0.46$) than for alternate forms ($d = 0.24$), was similar for tests of analytical ($d = 0.32$) and quantitative measures ($d = 0.30$), and, most importantly, the effects were much greater ($d = 0.70$) when there was some form of test coaching than when there was no such coaching ($d = 0.24$).

Coaching students for SAT tests has moderate effects on SAT performance, although the effects were greater on SAT mathematics than on verbal tests (Becker, 1990). Becker argued that the considerable variability in results of the examination of studies on coaching was because not all coaching is effective. Studies in which the coaching intervention included practice and instruction on answering particular items showed significant advantages over practice in taking complete examinations or instructions in general test-taking skills. The effects of coaching are greater when pre-tests are given in conjunction with the coaching program (Witt, 1993), and when the items in the test follow a format that is more complex and is not usually used (Powers, 1993).

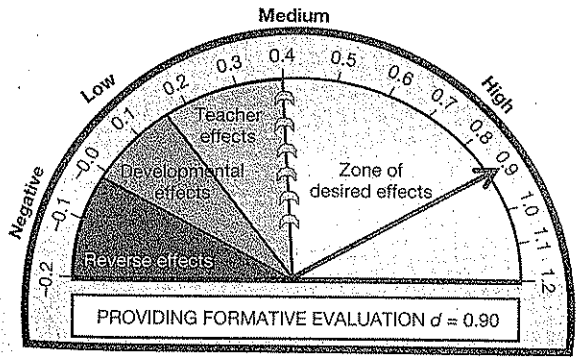
Another form of coaching is to become familiar with the examination process and examiner, particularly in one-to-one testing situations. In these situations, reducing anxiety about the testing context can make a difference. Fuchs and Fuchs (1985) found that examiner familiarity raised test performance by $d = 0.35$ standard deviations. Differential

performance favoring the familiar examiner condition was stronger when students were of low socioeconomic status, when students were tested on comparatively difficult tests, and when the examiner had been known to students for a longer duration. A further meta-analysis of the effects of examiner familiarity on student test performance (Fuchs & Fuchs, 1986) supported their 1985 findings. This meta-analysis also showed that students taking examinations scored higher when tested by familiar rather than unfamiliar examiners. The duration of the activity inducing familiarity had a strong positive influence on effect size. Again, low socioeconomic status students performed much better with a familiar examiner, while high socioeconomic status students performed similarly across examiner conditions.

Providing formative evaluation of programs

A major argument throughout this book is the power of feedback to teachers on what is happening in their classroom so that they can ascertain "How am I going?" in achieving the learning intentions they have set for their students, such that they can then decide "Where to next?" for the students. Formative evaluation provides one such form of feedback. Fuchs and Fuchs (1986) examined the effects of systematic formative evaluation by the teachers and found that this technique increased achievement for students with a mild learning disability ($d = 0.70$). The formative evaluations were effective across student age, treatment duration, frequency of measurement, and special needs status. When teachers were required to use data and evidence based models, effect sizes were higher than when data were evaluated by teacher judgment. In addition, when the data was graphed, effect sizes were higher than when data were simply recorded.

It is this feedback to teachers that assists in explaining why most of the more powerful effects are higher than what has been termed the "typical teacher effects" of $d = 0.25$ to $d = 0.40$. It is the attention to the purposes of innovations, the willingness to seek negative evidence (i.e., seeking evidence on where students are not doing well) to improve the teaching innovation, the keenness to see the effects on all students, and the openness to new experiences that make the difference. Interventions are not "change for change's sake" as not all interventions are successful. The major message is for teachers to pay attention to the formative effects of their teaching, as it is these attributes of seeking formative evaluation of the effects (intended and unintended) of their programs that makes for excellence in teaching.



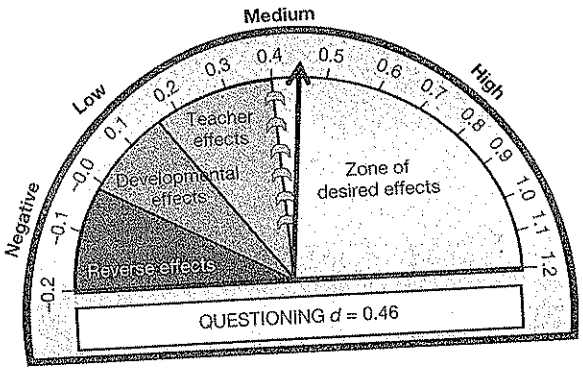
KEY	
Standard error	0.079 (Medium)
Rank	3rd
Number of meta-analyses	2
Number of studies	30
Number of effects	78
Number of people (1)	3,835

Questioning

Feedback can also come via teachers asking questions of their students, although it is an adage that teachers already know the answer to most of the questions they ask. The use of questions, especially higher-order questions, is often promulgated as a worthwhile teaching strategy: "Questioning opens up possibilities of meaning" (Gadamer, 1993, p. 375); "Questioning is a powerful strategy for building comprehension" (Mantione & Smead, 2003, p. 55); "Good questions lead to improved comprehension, learning, and memory of the materials among school children as well" (Craig *et al.*, 2006, p. 567). The study of the frequency, classification, and training of teacher questioning behaviors is based on the notion that skilled questioning by teachers can guide students to thoughtful and reflective answers and so facilitate higher levels of academic achievement (Samson, Strykowski, Weinstein, & Walberg, 1987).

So much of classroom time is spent with teachers questioning the students. Cotton (1989), for example, reviewed the evidence and found questioning was the second most dominant teaching method (after teacher talk), with teachers spending between 35–50 percent of teaching time posing questioning (e.g., Long & Sato, 1983; van Lier, 1998)—that is about 100 questions per hour (Mohr, 1998)—and the responses from the teacher to the students' answers to these questions was some form of judgment or correction, primarily reinforcing in nature, affirming, restating, and consolidating student responses. Brualdi (1998) claimed that teachers asked 300 to 400 questions per day, and the majority of these were low-level cognitive questions—60 percent recall facts and 20 percent are procedural in nature (Wilen, 1991) These are not open, inquiry questions, as students understand that the teacher already knows the answer (they are "display" questions; 82 percent are of this nature: Cotton, 1989). The reason for so much questioning relates to the conceptions of teaching and learning held by many teachers—that is, their role is to impart knowledge and information about a subject, and student learning is the acquisition of this information through processes of repetition, memorization, and recall: hence the need for much questioning to check that they have recalled this information.

The overall effects of questioning vary, and the major moderator is the type of question asked—surface questions can enhance surface knowing and higher-order questions can enhance deeper understanding. Samson, Strykowski, Weinstein, and Walberg (1987) used 14 studies to contrast the effects of predominantly higher cognitive questions and predominantly factual questions. Higher cognitive questioning strategies were found to have a small positive effect on learning measures. Factual pre-questions can facilitate learning



KEY	
Standard error	0.068 (Medium)
Rank	53rd
Number of meta-analyses	7
Number of studies	211
Number of effects	271
Number of people (0)	na

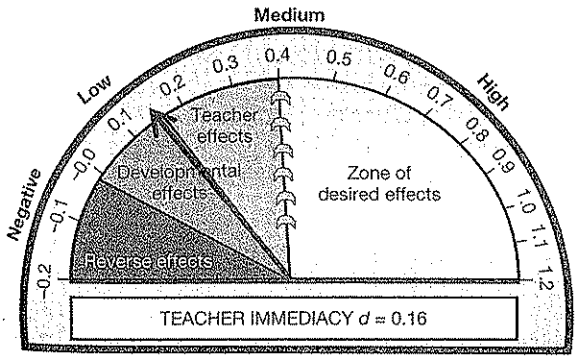
provided they are directly related to the texts or materials to be learnt (and have a negative effect when the questions asked are unrelated to the text material to be learnt, Hamaker, 1986). Higher-order questions are more effective on both direct and unrelated materials—"these results indicate that higher-order questions may have a somewhat broader general facilitative effect than factual adjunct questions" (Hamaker, 1986, p. 237).

Training in questioning matters. Gliessman, Pugh, Dowden, and Hutchins (1988) found that the questioning skills examined in the studies were very open to change through training. The general effect of training, academic level of trainees within training method, consistency of trainee certification level and pupils taught in practice, as well as consistency across practice and criterion teaching settings were all variables that had significant effects in the acquisition of questioning skills. Redfield and Rousseau (1981) also found that gains in achievement may be expected when teachers are trained in questioning skills. They found that lower level questions are more effective when aiming at surface level information, and a mixture of lower and higher level questions are more effective when aiming at deeper information and understanding. Studies designed to provide monitoring of program implementation show positive effects of 0.66 while those without monitoring showed negative effects (–0.10). Such attention by teachers to monitoring their own actions is powerful (and also reported in Gliessman *et al.*, 1988).

Perhaps of more importance than teacher questioning is analyzing the questions that students ask. As the work of my colleagues and I on the Socratic questioning in the Paideia project has demonstrated, structuring class sessions to entice, teach, and listen to students questioning of students is powerful (Hattie, *et al.*, 1998; Roberts & Billings, 1999).

Teacher immediacy

The immediacy and closeness of responses to the students shows them that teachers are listening and responding. "The applications of immediacy to educational settings introduced the idea that a teacher, through the use of certain cues, could reduce the perceived distance between instructor and learners and thereby influence certain classroom outcomes, especially student learning" (Allen, Witt, & Wheeless, 2006, p. 22). This immediacy is perceived by students as an acknowledgement of their engagement; it reduces the perceived distance between instructor and learners, is seen as rewarding to the student, and increases their level of enthusiasm or commitment to the learning task (Christophel & Gorham, 1995). The effects of teacher immediacy were much stronger on affective learning such



KEY	
Standard error	0.042 (Medium)
Rank	115th
Number of meta-analyses	1
Number of studies	16
Number of effects	16
Number of people (1)	5,437

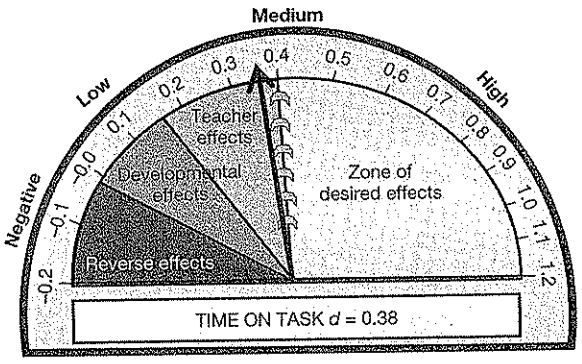
as attitudes to the teacher, course or engaging in the learning experience ($d = 1.15$) than on achievement ($d = 0.16$). From these results, and the correlation between affective and achievement learning, Allen *et al.* (2006) concluded that "teacher immediacy behaviors predict or cause a level of affective learning. In turn, the level of affective learning predicts or causes the level of cognitive thinking. … (the) teacher creates a motivational affective outcome that subsequently contributes to the generation of a cognitive outcome" (p. 26). They suggested that the teacher's immediacy also provided a source of feedback by the teacher about their interest, caring, and involvement in the student's learning.

Implementations that emphasize student perspectives in learning

The next set of topics relates to seeing learning from the student's perspective. Time on task, self-questioning, self-verbalization, peer tutoring, concept mapping, and the aptitude-treatment interaction.

Time on task

The typical claim is that practice makes perfect. I decided this was the case when I decided to play golf most mornings for a year. While my score dropped dramatically, there came a time when I realized that practice was not enough. Either professional coaching or a change to some physical predispositions would be needed. Further, we certainly do not want more time on task if the learning is not positive—it is like asking an unhealthy obese person to just eat more! Time on learning can involve: longer school days, longer school years, procedural time, time off-task, on-task time, and so on. There are various claims about how much actual time is spent in "engaged" learning time; Berliner (1984), for example, claims that about 40 percent of class time is spent on engaged time—and less of this engaged time is spent on productive time (which is that time that individual students find productive in their learning). So what happens in classes? Yair (2000) put wristwatches on 865 students (from 33 schools) that were programmed to emit signals (beeps) eight times a day for a week. When beeped, the students were asked to record what activity they were engaged in, and their thoughts and mood (which led to 28,193 daily experiences). The students were engaged with their lessons about half the total class time: engaged time was similar for boys and girls, but decreased over school grade. It was higher in mathematics than in English and social sciences, and was lowest when teachers were lecturing



KEY	
Standard error	0.101 (High)
Rank	70th
Number of meta-analyses	4
Number of studies	100
Number of effects	136
Number of people (0)	na

or when students were asked to watch television, and highest when students were working in groups or laboratories. The more students felt "challenged, and the greater the academic demand on students—the more the students are engaged with instruction—the less prone they are to external preoccupations" (Yair, 2000, p. 256).

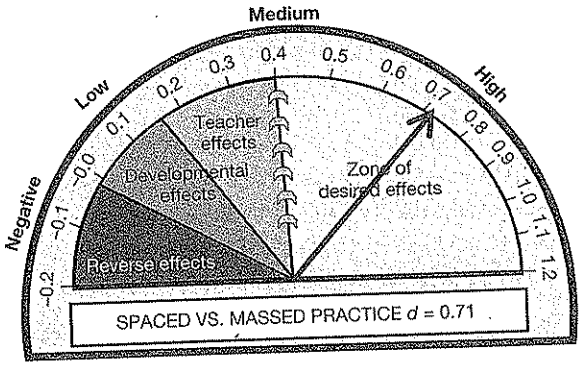
So at best half of student time in class involves engagement in the class activity—perhaps not surprising given so much time is spent listening (or pretending to listen) to teachers talking! Many have thus argued that making the available school time more productive should be the key to enhancing learning—and not merely extending the school day or year (Karweit, 1984; 1985): "Increasing allocated time, without increasing productive time, is unlikely to improve educational performance" (Walberg, Niemiec, & Frederick, 1994, pp. 98–99).

Fredrick (1980) explored the relationship between "engaged" instructional time and instructional outcome from 35 studies, and reported an effect size of $d = 0.34$. Lewis and Samuels (2003) found that more practice at reading was positively associated with reading ability, but the effect was only $d = 0.10$. The effects were slightly larger for grade 1–3 students, second language students, learning disabled students, and students reading below grade level: practice helps but it is not enough.

More important is that practice needs to be deliberate; particularly when first learning new material. Van Gog, Ericsson, Rikers, and Paas (2005) argued that it was not the amount of experience or practice in a domain that is relevant, but rather the amount of deliberate effort to improve performance. Deliberate practice refers to the relevant practice activities aimed to improve performance; it needs to be at "an appropriate, challenging level of difficulty, and enable successive refinement by allowing for repetition, giving room to make and correct errors, and providing informative feedback to the learner" (p. 75). Van Gog *et al.* further noted that such practice requires students to stretch themselves to higher levels of performance, and requires much concentration and effort over extended periods, usually of fixed times over many days. Feltz and Landers (1983) examined the effects of mental practice on motor-skill learning and performance and concluded that mental practice effects are found in both the initial and later stages of learning. Large effect sizes for cognitive tasks were more often achieved in a relatively short practice session and with only a few trials compared to motor and strength tasks.

Spaced and massed practice

It is the frequency of different opportunities rather than merely spending "more" time on task that makes the difference to learning. So teachers need to consider increasing the rate of correct academic responses to deliberative practice opportunities until minimal levels of mastery (defined by success criteria) are met (Walker, Greenwood, Hart, & Carta, 1994). This finding helps us to understand a common denominator to many of the effective practices in this book, such as direct instruction, peer-tutoring, mastery learning, and feedback. It is not over learning for the sake of it. Deliberative practice increases opportunities to not only enhance mastery but also fluency (the core of precision teaching). This is not "drill and practice", which so often can be: dull and repetitive; involve minimal feedback; not extend or provide multiple different experiences; not provide sufficient contextual variability to facilitate transfer of learning; and not be embedded in the context of the deeper and conceptual understandings that are part of the more total learning experience, and which so often aims at the surface knowledge. Deliberative practice can involve



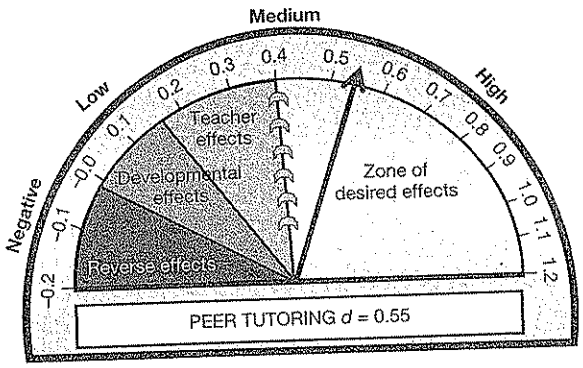
KEY	
Standard error	na
Rank	12th
Number of meta-analyses	2
Number of studies	63
Number of effects	112
Number of people (0)	na

specific skills and complex performances, and the attainment of success criteria can be motivating and certainly lead to longer retention of sometimes over-learned surface and deep knowing (Péladeau, Forget, & Gagné, 2003).

Nuthall (2005) claimed that students often needed three to four exposures to the learning—usually over several days—before there was a reasonable probability they would learn. This is consistent with the power of spaced rather than massed practice. Donovan and Radosevich (1998) concluded that students in spaced practice conditions performed higher than those in massed practice conditions ($d = 0.46$). Both acquisition ($d = 0.45$) and retention ($d = 0.51$) were enhanced by spaced rather than massed practice. The effectiveness of length of spacing was related to the complexity and challenge of the tasks—stronger effects were found for simple tasks with relatively brief rest periods, and longer rest periods were needed for more complex tasks (at least 24 hours or more).

Peer tutoring

The overall effects of the use of peers as co-teachers (of themselves and of others) in classes is, overall, quite powerful. If the aim is to teach students self-regulation and control over their own learning then they must move from being students to being teachers of themselves. One way to achieve this aim is to use peer tutoring—which too many consider a tool for older students to teach struggling younger children. While it is used for this purpose, the major influence is that it is an excellent method to teach students to become their own teachers. Reviews of tutoring literature have shown that peer tutoring has



KEY	
Standard error	0.103 (High)
Rank	36th
Number of meta-analyses	14
Number of studies	767
Number of effects	1,200
Number of people (3)	2,676

many academic and social benefits for both those tutoring *and* those being tutored (Cook, Scruggs, Mastropieri, & Casto, 1985). The overall effects from most of the meta-analyses on this topic are typically above the $d = 0.40$ average.

Hartley's (1977) meta-analysis of the effects on mathematics achievement of different instructional modes found that peer tutoring was the most effective of the various conditions she compared ($d = 0.60$). Peer tutoring was most effective when used as a supplement to, rather than a substitute for, the teacher roles. Cross-age tutors ($d = 0.79$) were more effective than same-age peers ($d = 0.52$) and adult tutors ($d = 0.54$). She also found a commonly reported conclusion: the effects on the tutors ($d = 0.58$) were not that different from the effects on those being tutored ($d = 0.63$) (see also Cook *et al.*, 1985, where supplemental was $d = 0.96$ and substitution was $d = 0.63$).

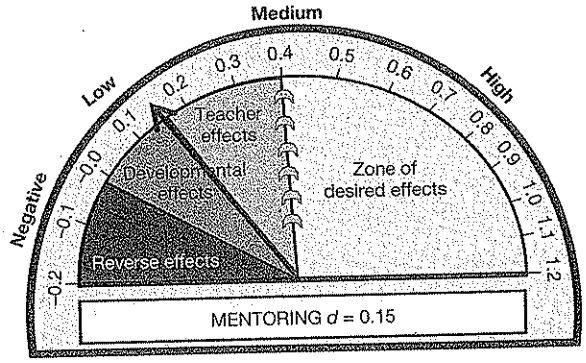
Peer tutoring has often been used with students with disabilities. Elbaum, Vaughn, Hughes, and Moody (2000) found that the magnitude of peer-tutoring effects did not differ according to whether disabled or non-disabled students acted as tutors or were doing the teaching. Cook, Scruggs, Mastropieri, and Casto (1985) reviewed studies where students with special needs were used as tutors of other students with special needs and found that those being tutored ($d = 0.53$) gained as much as those undertaking the tutoring ($d = 0.58$). Mathes and Fuchs (1991) found that peer tutoring was more effective than the instruction these students typically experienced. Kunsch, Jitendra, and Sood (2007) reported that these peer-mediated interventions were higher with disabled students in general ($d = 0.56$) than when they were in special classes ($d = 0.32$). Phillips (1983) found tutor methods were most effective with students in the acquisition rather than the proficiency phase of learning and when there were clear criterion measures (success criteria) used as targets.

Rohrbeck, Ginsberg-Block, Fantuzzo, and Miller (2003) found that peer interventions that were more student controlled (when peers are involved in setting goals, monitoring performance, evaluating performance, and selecting rewards), the effects were greater than when these were primarily controlled by teachers. When students were self-managers of their learning or the learning of others (in the peer-tutoring situation), then this autonomy led to greater achievement effects.

Thus, when students become teachers of others, they learn as much as those they are teaching. When they have some control or autonomy over this teaching, the effects are higher. It is likely that these effects are more critical when new surface level information is being taught, although it is likely that the tutors may need to understand the material at a deeper level to be effective teachers. This conjecture is not well explored in this literature and could well be subjected to further research. How often do we hear from teachers that "we learnt more when we were asked to teach it" but then see this maxim ignored as teachers enter classrooms and see students as recipients rather than producers of teaching and learning?

Mentoring

Mentoring is a form of peer tutoring, although it is normally involves older persons (often adults) providing academic or social assistance, or both, to younger people—but it also occurs throughout adult work situations to facilitate career development. Such mentoring assumes that supportive relationships with older people are important for personal, emotional, cognitive, and psychological growth. Mentoring usually involves little,



KEY	
Standard error	0.047 (Medium)
Rank	120th
Number of meta-analyses	2
Number of studies	74
Number of effects	74
Number of people (1)	10,250

if any, teaching and is more an "apprentice" model based on social and role model experiences. Mentoring had a close to zero effect on performance outcomes ($d = 0.08$), although there were higher effects on attitudes (satisfaction $d = 0.6$, school attitudes $d = 0.19$), and on motivation and involvement ($d = 0.11$) (Eby, Allen, Evans, Ng, & DuBois, 2008). That is, there is more change on attitudes than achievements, probably because "attitudes are more amenable to change than are outcomes that are more contextually-dependent" (p. 16). It was the case that effects were higher for academic mentoring than for youth (at risk, family-related mentoring) and workplace mentoring.

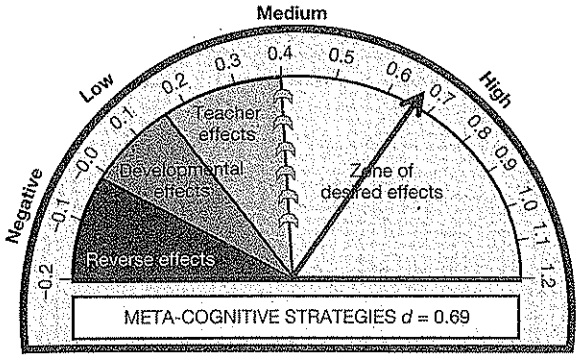
DuBois, Holloway, Valentine, and Cooper (2002) investigated many outcomes from mentoring. Across their 575 effect sizes, the average was $d = 0.18$ on achievement, and these low effects occurred when the program was one-on-one or in groups; the effects were lower in schools than in workplaces and higher for trained compared with non-trained mentors, but there was no relation with the frequency of contact nor the length of relationship between mentors and youth. The effects were similarly low for emotional or psychological outcomes ($d = 0.20$), problem and high risk behaviors ($d = 0.19$), social competence ($d = 0.16$), and career and employment outcomes ($d = 0.19$).

Implementations using student meta-cognitive and self-regulation learning

Meta-cognition relates to thinking about thinking. This section outlines a series of programs based on teaching various meta-cognitive strategies, including study skills, self-verbalization, self-questioning, aptitude-treatment interactions, matching learning styles, and individualized instruction.

Meta-cognitive strategies

Newell (1990) noted that there are two layers of problem solving: applying a strategy to the problem, and selecting and monitoring that strategy. Such "thinking about thinking" involved in this second layer of problem-solving has recently been referred to by the term "meta-cognition"; this refers to higher-order thinking which involves active control over the cognitive processes engaged in learning. Meta-cognitive activities can include planning how to approach a given learning task, evaluating progress, and monitoring comprehension. A synthesis of effective meta-cognitive training programs (Chiu, 1998),



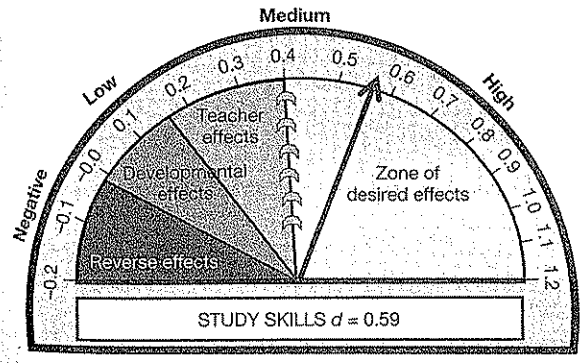
KEY	
Standard error	0.18 (High)
Rank	13th
Number of meta-analyses	2
Number of studies	63
Number of effects	143
Number of people (2)	5,028

found that such training is more effectively implemented using small-group instruction, with students in higher grades, with remedial students, and in less intensive programs. Haller, Child, and Walberg (1988) assessed the effects of meta-cognitive instruction on reading comprehension, and reported an effect size of $d = 0.71$ (see also Chiu, 1998). The most effective meta-cognitive strategies were awareness of textual inconsistency and the use of self-questioning. The more varied the instructional strategies throughout a lesson, the more students were influenced.

Study skills

Study skills interventions are programs that work on improving student learning using interventions outside what the teacher or teachers involved would normally undertake in the course of teaching. Interventions can be classified as *cognitive*, *meta-cognitive*, and *affective*. Cognitive interventions focus on the development of task-related skills, such as note taking and summarizing. Meta-cognitive interventions work on self-management learning skills such as planning; monitoring; and where, when, and how to use tactics and strategies. Affective interventions focus on non-cognitive features of learning such as motivation and self-concept (Hattie, Biggs, & Purdie, 1996). The argument in this section is that courses in study skills *alone* can have an effect on the surface level information, but it is necessary to combine the study skills *with the content* to have an effect on the deeper levels of understanding.

Lavery (2008) found a $d = 0.46$ effect on achievement from meta-cognitive study skills



KEY	
Standard error	0.090 (High)
Rank	25th
Number of meta-analyses	14
Number of studies	668
Number of effects	2,217
Number of people (8)	29,311

interventions. She found the highest effects from strategies that aimed at the "forethought" phase of learning; such as goal-setting and planning, self-instruction, and self-evaluation (Table 9.5). This strategy is "a major part of the forethought phase of this model (which occurs before the learner engages in the task) and has previously been shown to be a crucial aspect of interventions" (Greiner and Karoly, 1976, p. 497). Self-instruction occurs during the performance phase of the model and is an invaluable tool for guiding the learner through the focusing of attention and use of appropriate strategies. Self-evaluation concludes the cyclical model by allowing the learner to self-reflect on performance in

Table 9.5 Various meta-cognitive strategies and the effect sizes (Lavery, 2008)

Strategy	*Definition*	*Description*	*No. effects*	*ES*	*se*
Organizing and transforming	Overt or covert rearrangement of instructional materials to improve learning	Making an outline before writing a paper	89	0.85	0.04
Self-consequences	Student arrangement or imagination of rewards or punishment for success or failure	Putting off pleasurable events until work is completed	75	0.70	0.05
Self-instruction	Self-verbalizing the steps to complete a given task	Verbalizing steps in solving a mathematics problem	124	0.62	0.03
Self-evaluation	Setting standards and using them for self-judgment	Checking work before handing in to teacher	156	0.62	0.03
Help-seeking	Efforts to seek help from either a peer, teacher, or other adult	Using a study partner	62	0.60	0.05
Keeping records	Recording of information related to study tasks	Taking class notes	46	0.59	0.06
Rehearsing and memorizing	Memorization of material by overt or covert strategies	Writing a mathematics formula down until it is remembered	99	0.57	0.04
Goal-setting/ planning	Setting of educational goals or planning subgoals and planning for sequencing, timing, and completing activities related to those goals	Making lists to accomplish during studying	130	0.49	0.03
Reviewing records	Efforts to reread notes, tests, or textbooks to prepare for class or further testing	Reviewing class textbook before going to lecture	131	0.49	0.03
Self-monitoring	Observing and tracking one's own performance and outcomes, often recording them	Keeping records of study output	154	0.45	0.02
Task strategies	Analyzing tasks and identifying specific, advantageous methods for learning	Creating mnemonics to remember facts	154	0.45	0.03
Imagery	Creating or recalling vivid mental images to assist learning	Imagining the consequences of failing to study	6	0.44	0.09
Time management	Estimating and budgeting use of time	Scheduling daily studying and homework time	8	0.44	0.08
Environmental restructuring	Efforts to select or arrange the physical setting to make learning easier	Studying in a secluded place	4	0.22	0.09

relation to the previously set goals. While self-monitoring is very effective, it was not as high as that of self-evaluation, suggesting that self-monitoring in itself (such as ticking off completed tasks) can be much improved if taken a step further, where the learner actually evaluates what they have monitored.

The highest ranked strategy, that of organizing and transforming, has also been found to be a valuable component of many interventions (Hattie *et al.*, 1996). It is likely that the types of strategies included in this category (such as summarizing and paraphrasing) promote a more *active* approach to learning tasks. While several strategies such as record keeping, imagery, time management, and restructuring the learning environment were ranked lowest, it is likely that this is because they are more passive and involve non-active involvement with the content.

With regard to tertiary students, a closer examination of the effect sizes for these students shows that the smaller effects (and in one case a negative effect) generally came from the studies of shorter duration (i.e., those of a few days). Considering that the students in the tertiary studies were often identified as having difficulties with studying or were considered to be "at risk" by their institution, it seems that longer interventions may be required with these students. It is also likely that, as has been previously suggested, study habits are somewhat more "ingrained" with older students, thus making them more resistant to change (Hattie *et al.*, 1996). This was also indicated by one of the studies included in the meta-analysis, that of Nist and Simpson (1989), whereby achievement scores suffered an initial decrease after the implementation of the intervention, suggesting that a longer time frame is necessary, at least with tertiary-age students. There needs to be some un-learning of prior study skills before new learning can occur.

For students struggling to begin to understand, for lower achieving students, and for those wanting higher achievement, then teaching study skills can have advantages. Shrager and Mayer (1989), for example, claimed that note taking may facilitate better test performance for less skilled learners, but not for highly skilled learners. Mastropieri and Scruggs (1989) found the highest effect sizes of all for training special needs students with mnemonic methods of studying (see also Crismore, 1985; Kobayashi, 2005; Rolheiser-Bennett, 1986; Runyan, 1987)—although the effects of study skills programs for those struggling at the college level is quite low (Burley, 1994; Kulik, Kulik, & Shwalb, 1983). The mnemonic keyword strategies involve relating unfamiliar verbal stimuli into acoustically similar representations that become the keywords for remembering (e.g., Roy G. Biv for the colors of the rainbow). They did note that to maximize the chances of this knowledge being transferred and sustaining the learning, it was most effective when students were first able to *read* the text and determine what was important to remember, determine the optimal mnemonic strategy, correctly recall and implement the appropriate steps of strategy adaptation, and self-test, monitor, and correctly apply the learned information in the appropriate situation.

Kobayashi (2005) found that note taking effects were higher when students were given instructor's notes to work from ($d = 0.82$), as these provided exemplars for their own note taking and a rubric to work from when learning from the notes. The effects were higher when notes were provided ($d = 0.41$, compared to not provided ($d = 0.19$), and it was the reviewing of the notes ($d = 0.45$) that was more effective than the taking of the notes. He found no moderation effect relating to the length of the review, the presentation length that led to the taking of notes, or the format of the presentation (video, audio, or live).

Hattie, Biggs, and Purdie (1996) divided study skills programs into those aiming for

near- and far-transfer in terms of degree of transfer between training task and outcome measure, and whether they were more out of, or in-context of the discipline. They found greater effects of study skills programs on the lower order thinking tasks (e.g., memory, $d = 1.09$), than on reproductive performance in general ($d = 0.69$), and lower (but still high) on transformational performance ($d = 0.53$). As noted above, programs involving direct teaching of mostly mnemonic devices are highly effective with most students, and also conventional study skills training is effective for near transfer on low-cognitive-level tasks. Programs that were provided outside of the context of the subject matter (the more general study skills programs) were only effective when surface knowledge was the outcome, whereas programs run in-context (associated highly with the subject matter to be learnt) were most effective at surface and deeper knowing and understanding. We concluded that "the best results came when strategy training was used meta-cognitively, with appropriate motivational and contextual support" (Hattie *et al.*, 1996, p. 129) and questioned whether "learning-to-learn" programs that are not embedded in the context of the subject to be learnt are of much value. Three recommendations from the meta-analysis are that training should be (1) in context, (2) use tasks within the same domain as the target content, and (3) promote a high degree of learner activity and meta-cognitive awareness. "Strategy training should be seen as a balanced system in which the individual's abilities, insights, and sense of responsibility are brought into use, so that the strategies that are appropriate to the task at hand can be used" (Hattie *et al.*, 1996, p. 131). The student needs to know various strategies that are appropriate to the task at hand: the how, when, where, and why of their use. Strategy training needs to be embedded in the teaching context itself.

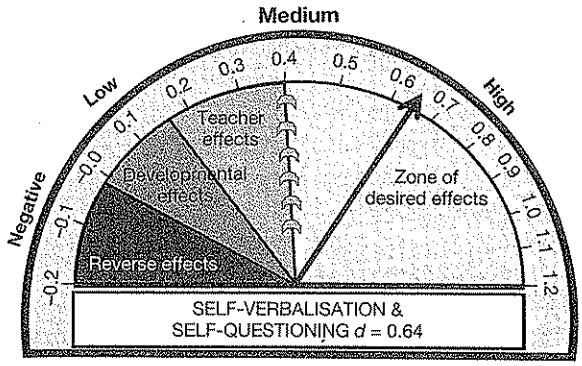
Study skills can also assist students to gain confidence that they are "learners" of the subject. Robbins, Lauver, Le, Davis, Langley, and Carlstrom (2004) found that the best study skills predictors of grade point average (GPA) were academic self-efficacy ($d = 0.38$), and that this confidence was as influential as high school GPA ($d = 0.41$), achievement motivation ($d = 0.26$), social involvement ($d = 0.12$), and academic goals ($d = 0.16$). Similarly, Ley and Young (2001) found self-efficacy to be among the best predictors of GPA ($d = 0.50$) and achievement motivation ($d = 0.30$), and that it had an incremental contribution over and above socioeconomic status, academic achievement, and high school GPA in predicting college outcomes. They argued that there were four principles to embed study regulation support in instruction:

1 guide learners to prepare and structure an effective learning environment;
2 organize instruction and activities to facilitate cognitive and meta-cognitive processes;
3 use instructional goals and feedback to present student monitoring opportunities;
4 provide learners with continuous evaluation information and occasions to self-evaluate.

These four principles can guide embedding study skills support in a wide variety of instructional media and contexts.

Self-verbalization and self-questioning

Self-questioning is one form of self-regulation, and given the comments in the previous section, are probably of more use to those in the early to intermediate phase of skill acquisition and for those of lower to middle ability (cf., de Bruin, Rikers, & Schmidt, 2007).



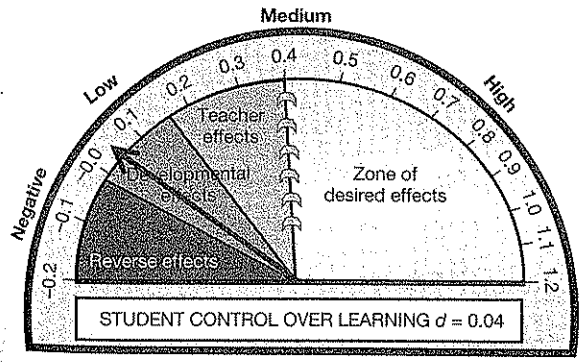
KEY	
Standard error	0.060 (Medium)
Rank	18th
Number of meta-analyses	3
Number of studies	113
Number of effects	1,150
Number of people (2)	3,098

Duzinski (1987) reviewed many procedures that taught a learning strategy or cognitive mediation strategy to students. Self-verbalization was among the most effective of the strategies, and it worked better for task oriented skills (e.g., writing or mathematics). In Huang's (1991) study of student self-questioning, the effects were higher with lower ability students. Similarly, Rock (1985) found that self-instructional training was effective for many students in special education programs.

Huang also noted that the use of self-questioning provided assistance in searching for the information needed, and thus increased student understanding of the messages of the material to be learned. Higher ability students were probably using a variety of self-regulation strategies already and self-questioning may not be as effective for them. The effects were higher for pre-lesson questioning ($d = 0.94$) and post-lesson questioning ($d = 0.86$), compared to questions interspersed during the lesson ($d = 0.52$); when the questionings were delayed ($d = 0.72$) compared to immediate ($d = 0.54$); and where there was teacher modeling ($d = 0.69$) compared to none ($d = 0.47$).

Student control over learning

The effect of student choice and control over learning is somewhat higher on motivation outcomes ($d = 0.30$) than on subsequent student learning ($d = 0.04$; Niemiec, Sikorski, & Walberg, 1996; Patall, Cooper, & Robinson, 2008). Indeed the more instructionally irrelevant choices had higher outcomes (e.g., color of pen to use, what music to listen to when



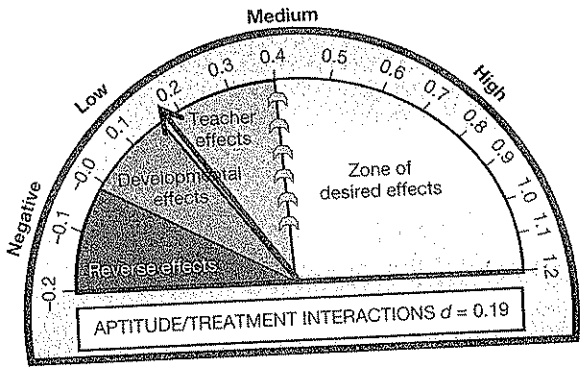
KEY	
Standard error	0.176 (High)
Rank	132nd
Number of meta-analyses	2
Number of studies	65
Number of effects	38
Number of people (0)	na

learning). Such irrelevant choices are less effortful and do not have major consequences on the learning, and too many choices may be overwhelming.

Aptitude-treatment interactions

There are many claims that instruction must be altered for different types of students. This very rich source of literature has commonly been identified by the term "aptitude-treatment interactions." There has been a long search for these aptitude-treatment interactions, and many researchers lost must interest after Cronbach and Snow (1977) produced a magnum opus on the subject. While they were optimistic that such interactions were critical and could be found, they still concluded that "well-substantiated findings regarding ATI [aptitude-treatment interactions] are scarce" (p. 6), and Glass (1970) claimed he did not "know of another statement that has been confirmed so many times by so many people" (p. 210). Since that time the search has continued, and many new aptitude-treatment interactions have emerged under headings such as learning styles (see next section), or differential treatments. All are premised on the search for instruction to accommodate individual differences.

There are few meta-analyses that provide evidence about aptitude-treatment interactions in general, possibly because most meta-analyses have been concerned with main effects. It is rare for meta-analyses to include information about interactions. Many include moderators (e.g., sex, age) but few include mediators, which are at the core of aptitude-treatment interactions (Cronbach & Snow, 1977). Whitener (1989) used the standardized interaction terms from 11 studies to find a weighted average regression coefficient—which is the best measure of the presence of an aptitude-treatment interaction. The average slope difference was about $d = 0.11$, and from her various careful analyses, she found support for the claim that students who have higher prior achievement benefit more than students with lower prior achievement from an increase in instructional support. That is, "higher achieving subjects capitalize on higher support, increasing the difference in performance between high and low achievers" (p. 78). It is important to appreciate that this effect of $d = 0.11$ is the effect after the main effects for prior achievement and treatment have been removed from the variance in learning—and this is worth considering (and as it is an aptitude-treatment interaction effect, it cannot to be compared to the other effects throughout this book). Pintrich, Cross, Kozma, and McKeachie (1986) claimed that aptitude-treatment interaction studies cannot be used with any confidence to construct general principles of



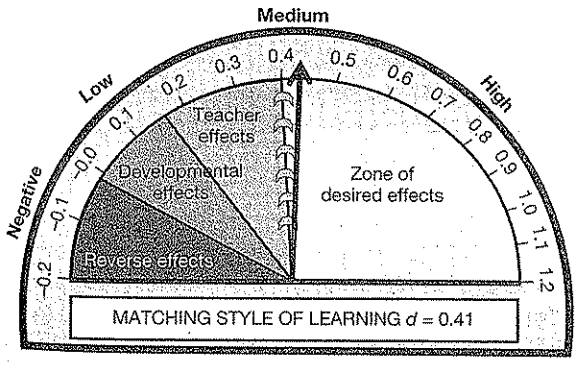
KEY	
Standard error	0.070 (Medium)
Rank	108th
Number of meta-analyses	2
Number of studies	61
Number of effects	340
Number of people (1)	1,434

instructional design, thus echoing Cronbach and Snow's (1977) earlier conclusion (based on a very comprehensive review of all possible research at that time) that "no Aptitude x Treatment interactions are so well confirmed that they can be used directly as guides to instruction" (p. 492).

Matching style of learning

Learning styles are one specific type of aptitude-treatment interaction and presume that different students have differing preferences for particular ways of learning. Often the claim is that when teaching is aligned with the preferred or dominant learning style then achievement is enhanced. For example, Dunn and colleagues (Dunn, Griggs, Olson, Beasley, & Gorman, 1995) claimed that students with strong learning styles, such as auditory, visual, tactile, or kinesthetic styles, showed greater academic gains as a result of congruent instructional interventions than those students who had mixed preferences or moderate preferences. Their model has five dimensions: biological (preference for warm vs. cool temperatures when learning), emotional (persistence vs. needing breaks when learning), sociological (working in groups or alone), physiological (intake while learning, mobility needs) and psychological (global versus analytic processing style differences). The claim is that teaching is more effective when these learning preferences are taken into account—although others have claimed the opposite: that we should be teaching students the learning styles they do not have (Apter, 2001).

It is hard to discern the meaning of some of these meta-analyses. One conclusion, given the average effect size of $d = 0.41$, is that learning style is somewhat important. But when we delve deeper, the model includes a mixture of attributes, especially the confusion of learning styles with learning strategies. Further, many of the meta-analyses correlate the learning style scores with achievement and thus are neither aptitude-treatment interactions nor learning style interventions. Many studies say no more than what students learn is correlated with achievement. Kavale and Forness (1987), for example, were interested in students with learning difficulties and found little support for the claim that there were higher outcomes when teaching students based on some supposed strength in auditory ($d = 0.18$), visual ($d = 0.09$), or kinesthetic $d = 0.18$) preference. Indeed they commented that "the groups seemingly differentiated on the basis of modality preferences actually revealed considerable overlap and it was doubtful whether any of the presumed preferences could really be deemed preferences" and "little (or no) gain in achievement was



KEY	
Standard error	0.016 (Low)
Rank	62nd
Number of meta-analyses	8
Number of studies	411
Number of effects	1,218
Number of people (6)	29,911

found when instructional methods were matched to preferred learning modality" (p. 237). Iliff (1994) found that no one style predicted achievement outcomes better than any other: $d = 0.28$ for diverger, $d = 0.29$ for assimilator, $d = 0.28$ for converger, and $d = 0.29$ for accommodator. He concluded that "since this study found the LSI [learning styles inventory] not to be a predictor of learner outcome and career fields of study, researchers will be advised to stop trying to fit square pegs into round holes" (p. 76).

Two meta-analyses seem so different from the others, and include so many errors that they should be excluded. Dunn, Griggs, Olson, Beasley, and Gorman's (1995) meta-analysis was mainly based on doctoral dissertations, many supervised by the authors, with mostly attitudinal outcomes, and many were based on adult samples. There are some unusual aspects in this meta-analysis. Some of the effects are large; Rowan (1988), for example, assigned teachers to in-service courses based on matched and mismatched learning style and preferences for time of day for instruction. The effect size reported is $d = 22.29$! This translates into a correlation between learning styles and achievement of $d = 0.996$—which is beyond the imaginable. The next largest correlation was $d = 0.887$ from Lashell (1986). She assigned 48 students to a control and 42 to a treatment group. Students' reading styles were evaluated and educational strategies recommended for each student (e.g., preferences were related to phonics-linguistics, whole-word, individualized, or language experience). Using a measure of reading as the outcome, Lashell used a regression analysis including grade, treatment or control group, gender, pre-reading score, teacher's years of education, and others. The Multiple R = 0.887 and Dunn *et al.* mistakenly used this R as the effect size—the pre-reading beta-weight, not surprisingly, is the largest predictor, and the treatment over control effect is relatively very small.

In many of the other studies in this meta-analysis there were similar problems; and some of the sample sizes were tiny. Zippert (1985) assigned nine adults to courses to match their (unspecified) learning styles and eight to a control course—both taught by the same instructor; the effect size was $d = 2.5$. Hutto (1982) asked four teachers to teach three classes where they were asked to match instruction to the students' learning preference and three where they were not so matched. Although a number of statistical tests were provided, only one was chosen to be interpreted—in third grade, the matched group exceeded the control group (and this is reported in the meta-analysis). Ingham (1989) gave 314 employees (route sales representatives, mechanics, and management) two lessons—one an auditory strategy with visuals, and one a tactual/kinesthetic strategy with visuals. When matched for preferences, there were differences in their attitudes towards the company training programs.

Overall in the Dunn *et al.* meta-analysis, the correlations were $r = 0.26$ for emotional, $r = 0.23$ for sociological, $r = 0.24$ for environmental, and $r = 0.46$ for physiological and outcomes. Given the studies in this latter group, it seems that matching learning to the students' preferred time of day for learning, intake preferences (food, snacking), mobile versus passive environments, and auditory preferences—but it is just not believable that the correlations of these effects exceed, in most cases, $r = 0.60$. For the same reasons, the meta-analysis by Sullivan (1993) should be disregarded. A student of Dunn, she synthesized 42 studies, but nearly all were the same as in the Dunn *et al.*, paper and included the same analysis flaws. Kavale, Hirshoren, and Forness (1998) also reviewed the Dunn *et al.* meta-analysis and concluded that the "weak rationale, curious procedures, significant omissions, and circumscribed interpretation should all serve as cautions" and that the study has "all the hallmarks of a desperate attempt to rescue a failed model of learning style" (p. 79).

It is difficult to contemplate that some of these single influences (such as whether you prefer to snack, or to sit up straight) explain more of the variance of achievement that so many of the other influences in this book. Mangino (2004), for example, noted that students enrolled in remedial courses had the highest achievement correlations with kinesthetic learning (doing, touching, interaction, $r = 0.64$), need for consistency in learning strategies and not learning in several ways ($r = 0.44$), a strong preference for intake (eating and drinking while learning, $r = 0.41$), and having an authority figure present when learning ($r = 0.34$). Higher achieving students had preferences for learning in several ways ($r = 0.31$), an authority-figure present ($r = 0.28$), the need for structure ($r = 0.38$), no sound ($r = 0.40$), a formal design (a preference to learn sitting up straight; back at a 90 degree angle, $r = 0.47$), and tended to be more motivated ($r = 0.25$). The message is that learners need teachers (authority figures), low cognitive load if in remedial classes, and multiple means of learning if in typical classes. The claims about need for snacking and sitting up straight defy my powers to make sense of them.

An alternative explanation is that when students enjoy learning then achievement is higher. The conditions under which they most enjoy learning are thus correlated, but it is the enjoyment of learning rather than the conditions that are critical. This would explain the correlations between various environmental influences and achievement. Lovelace (2005), for example, included a potpourri of studies relating achievement to modifying classroom environment, structured compared to unstructured situations, working alone or in pairs, effects of time of day of instruction, individual compared to other teaching methods. She argued that achievement is enhanced particularly when there is matching of preferences for mobility, light, auditory, tactual, or intake compared to matching on sound, temperature, design, or kinesthetic.

Slemmer (2002) was particularly interested in how technology-enhanced learning environments accommodate the learning styles of students. While she found small effects relating learning styles to outcomes, the highest effect was when the same treatment was provided for all students and not varying the instruction depending on learning preferences. Tamir (1985) related three cognitive preferences and learning and reported an effect size of $d = -0.28$ with recall (acceptance of information without consideration of implementations, applications, or limitations), $d = 0.32$ with principles (acceptance of information because it exemplifies or illuminates a fundamental principal, concept, or relation), $d = 0.24$ with critical questioning of information regarding its completeness, generalizability, or limitations, and $d = -0.06$ with application and emphasis on the usefulness and applicability of information in a general, social, or scientific context. Lower achievers prefer recall, whereas higher achievement is related to a preference for principles, critical questioning and application.

It is hard not to be skeptical about these learning preference claims. Holt, Denny, Capps, and de Vore (2005) asked whether teachers are able to perceive their students' learning preferences more accurately than random guessing. They found that the percentage correctly assessed was 30 percent whereas by chance the estimate was 25 percent—not a great show of confidence in teachers' ability to ascertain preferences. Coffield, Ecclestone, Moseley, and Hall (2004) completed an extensive analysis of various learning style models. There were few studies that met their minimum acceptability criteria, and they provided many criticisms of the field such as: too much overstatement; poor items and assessments; low validity and negligible impact on practice; and much of the advocacy in this is aimed at commercial ends. Learning strategies, yes; enjoying learning, yes; learning styles, no.

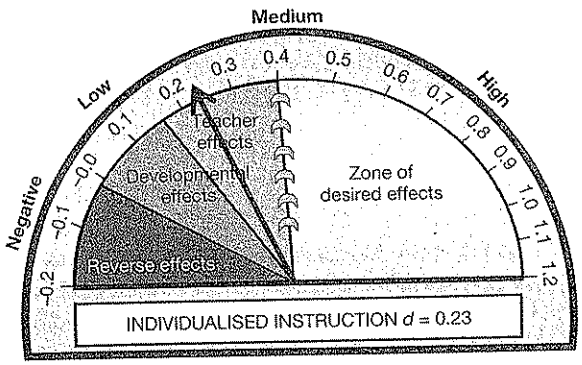
Individual instruction

Individualized instruction is based on the idea that each student has unique interests and past learning experiences, hence an individualized instructional program for each student allows for flexibility in teaching methods and motivational strategies to consider these individual differences. The evidence supporting individualized instruction, however, is not so supportive. Students are typically taught in classes of 20 or more; thus one of the major skills of teachers is to manage such classes, optimize peer co-teaching (even though this is not so common), and capitalize on the similarities and differences among the students.

Hartley's (1977) meta-analysis of the effects on mathematics achievement of different instructional modes found that individualized learning and programmed instruction were only slightly better than regular classroom instruction. In contrast, peer tutoring and computer-assisted instruction were more effective ($d = 0.60$) in increasing achievement. Similarly, Bangert, Kulik, and Kulik (1983) found that use of an individualized teaching system had only a small effect on student achievement in high school courses. There was limited contribution to student self-esteem, critical thinking ability, or attitude towards the subject matter taught when taught through individualized programs.

Waxman, Wang, Anderson, and Walberg (1985a, 1985b) claimed higher effects, but noted the importance of not just teaching the students by means of many individualized programs, but the importance of adapting instruction to the needs of students; ensuring these needs are based on the assessed capabilities of each student; using materials and procedures that allow students to make progress at their own pace; having periodic evaluations used to inform students about mastery; including aspects of self-responsibility for evaluating mastery; having student choice in educational goals; and aiming to have students assist each other in pursuing individual goals. There is no reason, however, why these attributes could not also occur in small or even larger groups.

Individualized instruction has been researched often in mathematics and science programs. Horak (1981) examined the effects of individualized instruction on mathematics achievement at elementary and high school level and found no significant difference to larger groupings. Similarly, Atash and Dawson (1986) examined the effects of the Intermediate Science Curriculum Study (ISCS), a semi-programmed, individualized course, and found that students on this program barely outperformed students taking a traditional junior high science curriculum ($d = 0.09$). Aiello and Wolfle's (1980) meta-analysis of individualized instruction in science in high school through college found individualized instruction to be similarly barely more effective than the traditional lecture approach ($d = 0.08$).



KEY	
Standard error	0.056 (Medium)
Rank	100th
Number of meta-analyses	9
Number of studies	600
Number of effects	1,146
Number of people (2)	9,380

Concluding comments

The argument defended in this chapter is that successful learning is a function of the worthwhileness and clarity of the learning intentions, the specifications, and the success criteria; the power of using multiple and appropriate teaching strategies with a particular emphasis on the presence of feedback focused at the right level of instruction (acquisition or proficiency); seeing learning and teaching from the students' perceptive; and placing reliance on teaching study skills and strategies of learning. Emphasizing learning styles, coaching for tests, mentoring, and individualized instruction are noted for their lack of impact.

The emphasis should be on what students can do, and then on students knowing what they are aiming to do, having multiple strategies for learning to do, and knowing when they have done it. It is teachers having teaching strategies aimed at enhancing the learning that was identified as the outcomes for the lesson, and who provide appropriate feedback to reduce the gap between where the student is and where they need to be. Both student and teacher need to set challenging goals, as this then sets the bar for the standards to be completed (at least, aiming for the h-point of 0.40 or higher effects), and to reach that bar challenging learning intentions, clear success criteria, and feedback will be needed. Setting challenging goals is a powerful part in the overall equation of what makes the difference in learning. Setting learning intentions invokes a "discrepancy-creative process", such that there is often a gap between present performance and where you wish to be (and which involves both teachers and students knowing where they are, where they are going, how they are going, what they need to do next, and how they can reduce this gap). Latham and Locke (2006), however, noted various pitfalls in goal setting, which highlight many of the factors of value noted in this chapter. When students *lack* the knowledge and skills to attain a goal, giving them a challenging goal sometimes leads to poorer performance than telling them to do their best. Goals may have an adverse effect on risk taking, if failure to attain a specific challenging goal is punished. Failures and false starts often are precursors to success. "Positive self-talk regarding an error ('I have made an error, great. I have learned something.') helps to keep our attention on the task rather than on ourselves ('How can I be so stupid?')" (p. 335).

The major messages in this chapter are the importance of learning intentions, success criteria, a classroom environment that not only tolerates but welcomes errors, attention to the challenge of the task, the presence of feedback to reduce the gaps, and a sense of satisfaction and further engagement and perseverance to succeed in the tasks of learning. This outline of successful teaching and learning is for all students—as another of my heroes, Sir Edmund Hillary, claimed with reference to himself, he was a man of modest abilities, and he combined these with a good deal of determination, and rather liked to succeed.