Journal of Experimental Psychology: General
2002, Vol. 131, No. 3, 349–363

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.3.349

Is Study Time Allocated Selectively to a Region of Proximal Learning?

Janet Metcalfe
Columbia University

Five experiments investigated whether people allocate their study time according to the discrepancy reduction model (i.e., to the most difficult items; J. Dunlosky & C. Hertzog, 1998) or to items in their own region of proximal learning. Consistent with the latter hypothesis, as more time was given, people shifted toward studying more difficult items. Experts, whether college students or Grade 6 children, devoted their time to items that were more difficult than did novices. However, in a multiple-trials experiment, people regressed toward easier items on Trial 2 rather than shifting to more difficult items, perhaps because Trial 1 feedback revealed poor learning of the easiest items. These findings are in opposition to the discrepancy reduction model and support the region of proximal learning hypothesis.

Although considerable progress has been made in understanding the characteristics of people's metacognitions as well as some of the mechanisms that underlie them (Dunlosky & Nelson, l992; Koriat, l993, l994, 2000; Koriat & Goldsmith, l996; Metcalfe, l993a, l993b, 1996, l998, 2000; Metcalfe, Schwartz, & Joaquim, l993; Miner & Reder, l994; Nelson & Narens, l994; Reder, l988; Reder & Ritter, l992; Schwartz & Metcalfe, l992, l994; Schwartz & Smith, l997; Thiede, l996; Thiede & Dunlosky, l999), our knowledge of how these metacognitive processes influence memory is less secure. Presumably, one of the central reasons that self-reflective knowing is important is because metacognitions should allow people to exert control over their own future learning and memory (see Koriat, 2000; Nelson, 1990; Nelson & Narens, l994). One straightforward control process that appears to be guided by people's metacognitions is the allocation of study time.

The empirical work on study-time allocation, until very recently, seemed entirely consistent. In an exhaustive review of the literature, Son and Metcalfe (2000) found 19 published studies concerned with how people allocate study time, with a total of 46 treatment combinations (consisting, for example, of different experiments, age groups, participant populations, or materials). The consensus in the literature was overwhelming. Thirty-five of these treatment combinations showed that people allocated more study time to items that were judged to be more difficult (or which by some objective measure were more difficult) rather than to the easy items.[1] Three treatment combinations showed that extra time was allocated to items of intermediate difficulty.[2] The other eight conditions showed null results with no discernible correlation between difficulty and study-time allocation. The data for these null results came from mildly retarded teenagers (Belmont & Butterfield, l971) and young children (Bisanz et al., l978; Dufresne & Kobasigawa, l988, l989; Kobasigawa & Dufresne, l992; Masur et al., l973) and have been interpreted as indicating that these individuals may have failed to devote more time to the more difficult items because (a) they were unable to adequately assess the difficulty or (b) they were unable to apply appropriate study strategies as a function of their metacognitions. None of the 46 treatment combinations showed that people had a preference to allocate their study time to the judged-easy items. The literature points to a simple conclusion: As long as they are metacognitively adept, people allocate their time to the items they judge to be difficult. As Mazzoni and Cornoldi (l993) noted, "the most logical strategy is to give more attention (i.e., allocate a greater proportion of available resources) to difficult items and less attention to easier ones" (p. 47).

These data have a natural explanation in terms of the discrepancy reduction model of study-time allocation (Dunlosky & Hertzog, 1998), in which people are thought first to determine, for each item, its degree of discrepancy from the desired state of learning. Then they devote the most study time to those items with the largest discrepancies. There is now only one study (Son & Metcalfe, 2000) that casts doubt on the generality of both the data and this model. However, despite the plausibility of the strategy given in the discrepancy reduction model as an explanation of the above-cited literature, it is likely to be a maladaptive learning strategy. Indeed, the small amount of data bearing on this issue suggests that allocating additional study time to difficult items does not necessarily result in performance gains. When Nelson and Leonesio (l988) investigated the relation between freely allocated study time, judgments of learning (JOLs), and performance, emphasizing either speed or accuracy, people in the accuracy condition spent up to 10 s more per item than when they were in the speeded condition. They tended to devote this extra time particularly to the more difficult items. But there was very little gain in performance.

I thank the James S. McDonnell Foundation for supporting this research. I thank Nate Kornell for his many contributions. I also appreciate the help of Lisa Son, Dan Kimball, and Janina Pietrzak.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, New York, New York 10027. E-mail: metcalfe@columbia.edu

[1] Belmont & Butterfield, l971; Bisanz, Vesonder, & Voss, l978; Cull & Zechmeister, l994; Dufresne & Kobasigawa, l988, l989; Dunlosky & Connor, l997; Dunlosky & Hertzog, l997; Kellas & Butterfield, l971; Kobasigawa & Dufresne, l992; Kobasigawa & Metcalf-Haggert, l993; Le Ny, Denhiere, & Le Taillanter, l972; Masur, McIntyre, & Flavell, l973; Mazzoni & Cornoldi, l993; Mazzoni, Cornoldi, & Marchitelli, l990; Mazzoni, Cornoldi, Tomat, & Vecchi, l997; Nelson, Dunlosky, Graf, & Narens, l994; Nelson & Leonesio, l988; Thiede & Dunlosky, 1999; Zacks, l969.

[2] Mazzoni & Cornoldi, 1993; Mazzoni et al, 1990.

Indeed, the lack of an effect was so apparent that the authors dubbed their results the “labor in vain” effect (p. 681).

Furthermore, thinkers as diverse and distinguished as Hebb (l949), Piaget (l952), Vygotsky (l987), Berlyne (l978), and Atkinson (l972a, 1972b) have all argued that there is a region of materials or concepts just beyond the grasp of the learner that is most amenable to learning. The learner’s efforts, according to these views, should be most effective in what I here call a *region of proximal learning* (following Vygotsky’s, 1987, zone of proximal development and corresponding roughly to what Atkinson called the “T” or transitional state—a state in which the items are neither fully learned nor completely unlearned). The predictions concerning what should be studied according to the discrepancy reduction model (those items with the greatest difference from the desired learned state) and what should be studied according to the region of proximal learning hypothesis (those items with the smallest distance from being learned) are in opposition.

It is possible, of course, that a person’s region of proximal learning might sometimes correspond to the most difficult items, if the person were very accomplished (i.e., an expert), or the materials were constructed to that end, or if the person had an unlimited amount of time, or the materials, overall, were rather easy. Thus, there may be some particular cases in which the discrepancy reduction model and the region of proximal learning hypothesis overlap. However, otherwise, if the region of proximal learning hypothesis is correct, people’s adherence to the particular selective study-time-allocation policy of devoting their time to the most difficult items should only occur because their control strategies were faulty: They either did not know or could not use the knowledge that their best chance at learning came from devoting time to items in their own proximal region.

Atkinson (l972a) has been very specific on this point of people’s purported metacognitive and control inepitude:

> One way to avoid the challenge and responsibility of developing a theory of instruction is to adopt the view that the learner is the best judge of what to study, when to study, and how to study. I am alarmed by the number of individuals who advocate this position despite a great deal of negative evidence. My data, and the data of others, indicate that the learner is not a particularly effective decision maker. (p. 930)

He showed that his computer-based study program, in which he computed what he thought people should study and allocated participants’ study time to those materials selectively, resulted in better performance than did either people’s own choice of study items (following his instructions on what to avoid) or a computer-controlled study program that gave equal time to all items. These results suggest that study time devoted to the appropriate items may not be in vain but that people may not normally choose the appropriate items.

Although the view shared by Piaget (1952), Vygotsky (1987), Berlyne (1978), Hebb (1949), and Atkinson (1972a, 1972b) seems reasonable as a learning theory and as a strategy people should employ if they are to reap the maximum gain from their efforts, none of the experiments cited by Son and Metcalfe (2000) showed that people devoted their time to the easiest items, as surely should be true in some cases. These seemingly unassailable results, though, may not be quite as solid as the overall summary statistics suggest (35 of 46 studies supporting the discrepancy reduction theory with a few ties or null results but no reversals). It is notable, for instance, that Mazzoni’s three studies (Mazzoni & Cornoldi, 1993; Mazzoni et al., 1990, 1997) did show that sometimes people devoted their time to items of intermediate difficulty, and the Thiede and Dunlosky (1999) study showed that people sought out the easier items first—both supportive of the region of proximal learning hypothesis. Furthermore, many of the studies in the literature are close variants of one another, and so if the difficult items happened to be the proximal items by accident in some early study, that accident might have been propagated through the literature and might suggest more concordance with the discrepancy reduction model than is warranted. In addition, nearly all of the studies presented the items one at a time, and people were allowed to study them for an unlimited amount of time. This particular experimental paradigm—offering neither choice nor time constraints—may have biased the results toward a particular conclusion. Finally, the predictions of the proximal view tend to be nonmonotonic—that people will benefit most by studying in the mid-range—not too easy and not too hard. However, unpredicted nonmonotonic patterns might well show up as either zero correlations or as null results under the assumptions of linearity usually used. So the seemingly unassailable data favoring the discrepancy reduction model might not be so immune to challenge as one might originally suppose.

Further, although the literature reviewed by Son and Metcalfe (2000) favored the discrepancy reduction model, their own findings (also see Thiede & Dunlosky, l999) went against it, providing evidence that people may sometimes choose to study the easy rather than the difficult materials. Although their results were suggestive, they were not definitive. They found this effect only when the materials were long passages—pages of biographical text materials, for example, or Elizabethan sonnets—and when the participants were extremely time pressured. With shorter materials—such as haiku—people devoted their study time to those materials judged to be more difficult. Thus, one cannot ascertain whether the selective study of easy materials that they found might have been a reading effect that obtained only with long passages or if it was really related to the intrinsic learnability of the materials, as the region of proximal learning hypothesis suggests. In addition, in each of their three experiments, Son and Metcalfe asked participants to provide judgments of interest as well as judgments of difficulty, and in each there was a positive correlation between judged interest and allocation of study time. Thus, Son and Metcalfe’s study-time-allocation results might have been based on interest, rather than on assessments of difficulty.

The region of proximal learning hypothesis predicts that variables such as item difficulty, expertise of the participant, number of trials spent studying, or total study time available should affect the selection of items to which time is allocated and the susceptibility of the items to learning gains. People’s shifts toward more challenging materials provide the underpinnings for developmental theories such as those of Piaget (1952) and Vygotsky (1987). Suppose easy, medium, and difficult items are intermixed. Initially, before anything has been learned, the easy items will be in the proximal learning state and will gain most from study. Once the easy items are mastered, though, little additional gain would be expected for additional study effort on them, and the region of proximal learning should shift to an item set that is more difficult. Learning, then, is reflected in a shift toward study of items of progressively greater difficulty. Allowing additional study time or additional study trials should produce a similar shift in the region of proximal learning toward the more difficult items. Contrasts between novices and experts should also show the same pattern. The region of proximal learning for the novice should be of lower difficulty than is that of the expert. These predictions were tested in the experiments presented here.

Experiment 1

In this experiment, the materials were single-word translations from English into Spanish. Thus, the question of whether or not selective allocation of study time to the easy items can occur with materials other than long texts (where it might be an issue of reading rather than of learning) could be assessed. On the basis of a pilot experiment, the difficulty of the items was rather cleanly separated into items that were of easy, medium, and difficult levels, and these objective levels of difficulty, from the same prior study, accorded well with people's assessments. In that experiment, we had constructed 200 Spanish–English word pairs of vocabulary translations, 50 of which were identical in the two languages, 50 of which were very easy to learn, 50 of which were of moderate learning difficulty, and 50 of which were extremely difficult to learn. Participants, who were Columbia University students from the same population as in the present studies, first saw an English word, randomly selected, and then were asked for their response in Spanish. The computer then gave them the Spanish word and asked for their judgment of whether that pair was in one of three states: learned (L), transitional (T), or unlearned (U). In an earlier experiment, both numerical JOLs and L–T–U judgments had been requested, and we had found that the two were highly related to one another, and, if anything, the predictive value of the L–T–U judgments was slightly, but not significantly, better than that of the numerical JOLs. The participants then went through certain learning procedures that are not detailed here. The JOLs (L–T–U judgments), after the one very brief exposure, were highly related to item difficulty, such that people gave mainly L judgments for identical and easy items (L, T, and U judgments were 86%, 7%, and 7%, respectively, for the identical items and 67%, 23%, and 10%, for the easy items), T and U judgments dominated for the medium items (23%, 35%, and 42%, for L, T, and U, respectively), and primarily U judgments were given for the difficult items (7%, 16%, and 77%, for L, T, and U, respectively). Thus, people appeared to have a good idea of whether items were learned, transitory, or unlearned, and these values corresponded very well both to more standard JOLs to the a priori categories used and to their eventual performance. We used these same materials (excluding the identical items since they seemed to be characterized by a different learning process than the other items) in the experiments that followed, with the confidence that we did not have to take the judgments initially since they were highly reliable and our categories were quite well defined.

The pilot study had also shown that when people had too many choice alternatives (20, in the case of that study), they did not make real strategic choices among the items. Memory load and complex choice factors resulted in their resorting to strategies such as processing in serial order. Therefore, to encourage the subjects to choose freely—and not to be overly controlled by either the memory limitations of having to recollect which items were easy and difficult or by the choice situation itself when a large number of items was simultaneously presented—only three item pairs were presented at any given time in the current experiment. One of the item pairs was easy, one moderate, and one difficult. As well as varying the difficulty of the pairs, the amount of time people were allowed to study each triad of pairs was varied. They had either 5 s,15 s, or an unlimited time (up to 1 min) for each triad.

The predictions of the region of proximal learning hypothesis were that when little time was allowed, people would devote the bulk of their time to the easy items, less time to the medium items, and less time still to the difficult items. As they had more time available, however, there would be a shift toward devoting more and more time to the more difficult items because the easy items would have been learned, leaving the additional time free to devote to harder, as-yet-unlearned items.

Method

Participants

Participants were 12 Columbia University students who received partial course credit for their participation. Each participant was tested individually on an iMac computer. Participants were treated in accordance with the ethical standards of the American Psychological Association.

Design

The experiment was a 3 (difficulty of materials: easy, medium, difficult) × 3 (time allowed per triad: 5 s, 15 s, or unlimited, up to 1 min) within-participants design with 16 pairs per treatment combination. The order of the timing conditions, which was a blocked variable, was counterbalanced across participants.

Materials

The materials were 144 English–Spanish vocabulary pairs that were divided into three sets of 48 in each of the three (easy, medium, and difficult) conditions. Appendix A provides a listing of all 144 word pairs.

Procedure

Participants were informed that they would be asked to learn 144 English–Spanish vocabulary pairs. They were instructed that on each trial, three English words would be presented on the computer screen and that the Spanish translation of a given word would be displayed when they clicked on the question mark button below that word (and that the Spanish translation would remain on the screen until they clicked on another question mark or the study time elapsed). Participants were informed that the items on the left were relatively easy, those on the right relatively difficult, and those in the middle intermediate and that they were free to allocate their study time however they wished. They were also informed of the study time allowed for each triad in each block of the study phase. During the 5-s and 15-s blocks, the computer automatically proceeded to the next trial when the time period elapsed, whereas in the unlimited-time block a "next triad" button was presented along with each triad, and participants clicked it when they wanted to begin studying the next triad.

When the person had completed the study phase, the computer administered a test by displaying an English word and asking the participant to type the Spanish translation and then hit the Return button for the next English word. Participants were told that they could change their answers up until they hit the Return button and that their data would be scored both strictly and leniently, indicating that if they had some idea of what the words might be that they should type in their answers, even if they were not entirely certain.

Results

In the analysis of variance data presented here, a cutoff value of $p < .05$ was chosen as the criterion for significance. Tukey tests were used throughout as post hoc analyses on simple effects. Post hoc *t* tests were computed on individual comparisons when interaction terms were significant, and the *p* values for these are reported directly since there is no generally agreed upon authoritative reference for what the alpha level should be. The recall data were computer scored in two ways: strictly (where the item had to be perfect, with no spelling mistakes or alterations of any sort to attain credit, and each item was assigned either a 0 or 1 value, resulting in a percentage score for each participant when his or her data were collapsed over the 16 observations in each treatment combination) and leniently. The algorithm for lenient scoring, which were written by Brady Butterfield, assigns a value between 0.0 and 1.00 to the answer, on the basis of the extent of letter overlap between the response and the correct answer. Generally, items with scores over about .75 tended to be what a human scorer would call spelling mistakes. Items with scores over .75 were assigned a 1.0, items with scores of .75 or less were assigned a 0.0, and then the scores were averaged over the 16 trials. Data from both analyses are presented. The study-time-allocation data were also analyzed in terms of the proportion of time spent on each item type.

Recall Performance

As expected, with strict scoring, there was a main effect of item difficulty, $F(2, 22) = 57.64$, $MSE = .03$, effect size (ES) = .84. Tukey tests showed that there were significant differences between the easy and medium difficulty items, the easy and difficult items, and the medium and difficult items. There was an effect of time allowed on performance, $F(2, 22) = 9.787$, $MSE = .02$, ES = .47. More time resulted in better performance. The interaction between time allowed and difficulty was significant, $F(4, 44) = 6.20$, $MSE = .01$, ES = .36, as shown in Table 1. Study time tended not to affect the most difficult items, perhaps because of a floor effect with those items. This interaction, though, was not significant when the data were scored leniently (shown in Figure 1A), and so is not interpreted further. With lenient scoring, the effect of difficulty, $F(2, 22) = 119.07$, $MSE = .03$, ES = .91, and time allowed, $F(2, 22) = 15.40$, $MSE = .02$, ES = .58, were both significant, as with strict scoring. Tukey tests showed that with both strict and lenient scoring, people did better in the easy condition than in the medium condition, and they did better in the medium condition than in the difficult condition.

Table 1
Strictly Scored Recall Performance for Experiments 1–4 (and 95% Confidence Intervals)

	Time allowed		
Item difficulty	5 s	15 s	Unlimited
Experiment 1: Novice adults			
Easy	0.292 (0.083)	0.505 (0.149)	0.542 (0.144)
Medium	0.068 (0.041)	0.109 (0.061)	0.198 (0.104)
Difficult	0.000 (0.000)	0.005 (0.010)	0.031 (0.032)
Experiment 2: Expert adults			
Easy	0.722 (0.136)	0.778 (0.082)	0.840 (0.079)
Medium	0.285 (0.058)	0.465 (0.154)	0.569 (0.103)
Difficult	0.097 (0.068)	0.188 (0.119)	0.312 (0.157)
Experiment 3: Novice children			
Easy	0.025 (0.033)	0.219 (0.094)	0.219 (0.069)
Medium	0.006 (0.012)	0.013 (0.016)	0.031 (0.033)
Difficult	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Experiment 4: Expert children			
Easy	0.449 (0.096)	0.455 (0.095)	0.455 (0.129)
Medium	0.091 (0.038)	0.091 (0.048)	0.148 (0.080)
Difficult	0.011 (0.015)	0.028 (0.034)	0.011 (0.022)

Study-Time Allocation

When the absolute time was analyzed, of course, people allocated more time in the conditions in which more time was allowed them. Because there was always a pause before they began to study the first item in every triad, however, the total time means were not 5 s and 15 s in those two respective conditions but rather were 3.01 s and 11.16 s. In the unlimited time condition, the total time spent studying was 33.81 s. There was a main effect of difficulty, $F(2, 22) = 9.21$, $MSE = 9.60$, ES = .46, and an interaction between time allowed and difficulty, $F(4, 44) = 6.27$, $MSE = 6.85$, ES = .36, showing that as more time was allowed, people's study time shifted toward the more difficult items. At the 5-s rate, the difference between time allocated to easy and difficult items favored the easy items, $t(11) = 2.43$, $p = .03$; at the 15-s rate, there was no difference between time allocated to easy and to difficult items, $t(11) = -1.14$, $p = .28$; and at the unlimited-time rate, there was a reversal such that more time was allocated to the difficult than to the easy items, $t(11) = -2.1$, $p = .05$.

The same pattern emerged when the amount of time spent studying was analyzed proportionately rather than in terms of raw times. As the post hoc Tukey tests confirmed, people tended to allocate proportionately more time to the medium items than to either the easy or the difficult items, overall, $F(2, 22) = 7.33$, $MSE = .04$, ES = .40. However, as the interaction between time allowed and difficulty indicates, $F(4, 44) = 5.89$, $MSE = .03$, ES = .35, and as Figure 1B shows, when people had little time, they devoted it to the easy and medium difficulty items; when they had unlimited time, there was a shift toward studying the more difficult items, as was predicted by the region of proximal learning hypothesis. At the 5-s rate, the difference between time allocated to easy and difficult items favored the easy items, $t(11) = 2.79$, $p = .02$; at the 15-s rate, there was no difference between time allocated to easy and to difficult items, $t(11) = -1.08$, $p = .30$; and at the unlimited-time rate, there was a reversal such that more time was allocated to the difficult than to the easy items, $t(11) = -2..41$, $p = .03$.

Discussion

In contrast to the predictions of the discrepancy reduction model, people did not selectively or exclusively devote their study time to items that were very difficult. Rather, they devoted it selectively to the medium-difficulty items, overall. This result is consistent with the idea that people were attempting to study items in their own region of proximal learning. Of even more interest was the finding that which items people chose to study depended on how much total time they had for each triad. According to the region of proximal learning hypothesis, as the amount of study time available increases, so too does the difficulty of items that can

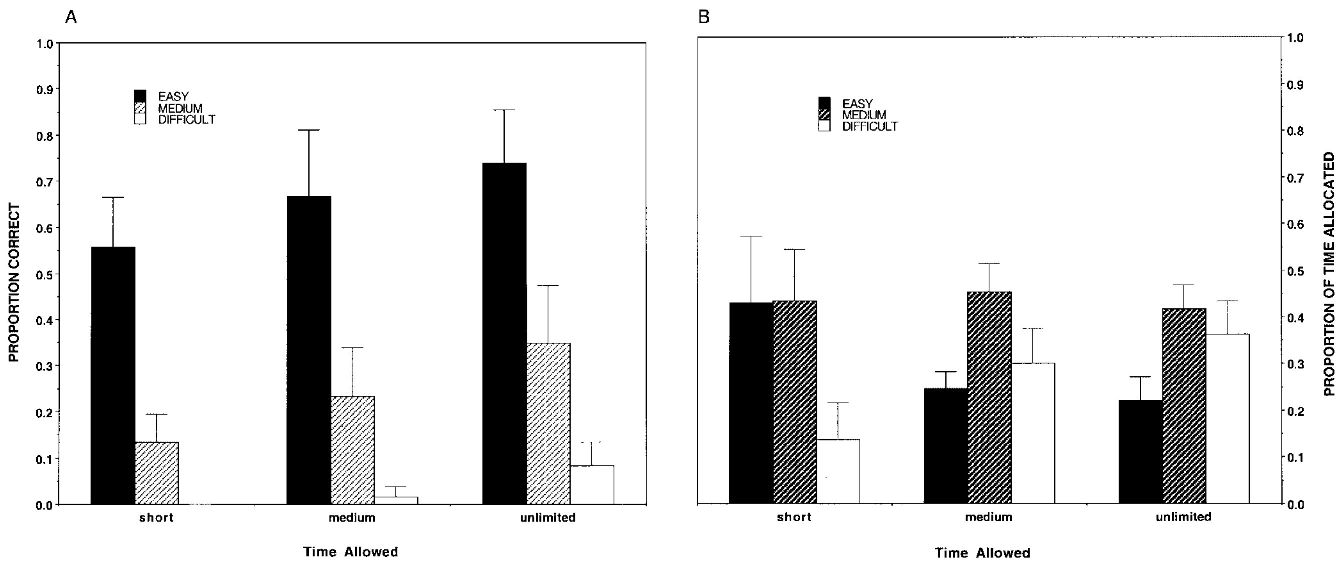

Figure 1. Experiment 1. A: Recall, scored leniently, of easy, medium, and difficult items. B: Proportionate time devoted to easy, medium, and difficult items. Error bars indicate 95% confidence intervals.

be profitably studied in the allowed time. The data in this experiment were very systematic in showing this predicted shift with increased study time.

Experiment 2

This experiment investigated the idea that the region of proximal learning is different for different people and depends critically on what the person already knows. It is notable that, in the past literature, the few exceptions to the general pattern that people have devoted their study time to the difficult items have come from studies with young children and mildly retarded teenagers. These null results—in which people apparently did not devote their time to the most difficult items—were interpreted as indicating that those groups had faulty metacognitions or were unable to allocate their study time appropriately. Such lack of metacognitive sophistication might even be thought to impair performance, and, indeed, a random pattern of study-time allocation, or a lack of strategy, might indeed impair performance. However, the region of proximal learning hypothesis suggests an alternative interpretation. It may be that the region that was optimal for study was different for the young children than it was for the older children. If these younger children were appropriately monitoring their own region of proximal learning and if they were using good rather than poor control strategies, they would have devoted their time to the items that were easier. Thus, these apparently null findings might have indicated good metacognitive monitoring and control, rather than the reverse, and support the region of proximal learning hypothesis. The present experiment, which uses adult experts, as well as Experiments 3 and 4 in which children who are either novices or experts were the participants, investigate this possibility.

To determine directly whether study-time allocation varied as a function of a person's expertise, the participants in this experiment were Spanish experts. The prediction of the region of proximal learning hypothesis tested in Experiment 2 was that people who have a greater degree of learning in English–Spanish vocabulary would choose items that are more difficult than did the novices—their region of proximal learning should be systematically shifted to the right.

Method

The experiment was identical to Experiment 1 except for the expertise of the participants, who were, as before, from the introductory psychology class pool at Columbia University. However, in this experiment, all participants had declared themselves to be experts in Spanish, whether because Spanish was their native tongue or because they had spent more than 4 years studying the language and claimed to be expert. In fact, though, of the 17 self-proclaimed experts, only 9 attained a performance scored strictly higher than 30% correct overall. In the basic analyses presented next, only data from those 9 people are included. Insofar as this experiment was conducted at the same time as was Experiment 1 (aside from 3 participants who were tested at a latter time), and inclusion in Experiment 1 or Experiment 2 depended only on the participant's knowledge of Spanish, and the experiments were in every other way identical, they were compared statistically. The self-proclaimed experts who scored less than 30% correct were analyzed as a separate group and compared both to the true experts in Experiment 2 and to the novices from Experiment 1. The comparisons with the 8 "pseudoexperts" are shown in Appendix B.

Results

Recall Performance

As can be seen from Table 1 and from Figure 2A, the experts in this experiment, like the novices in the previous experiment, found the easy items to be easy, the medium items to be of more difficulty, and the difficult items to be difficult. The effects of difficulty assessed with lenient scoring, as well as with strict scoring, were significant, $F(2, 16) = 90.23$, $MSE = .03$, ES = .92, and, $F(2, 16) = 143.50$, $MSE = .02$, ES = .95, respectively. Tukey tests showed significant differences in accuracy between easy and medium, medium and difficult, and easy and difficult items—both for lenient and for strict scoring. Experts' performance on the easy items approached ceiling. (One reason that it was not absolutely on

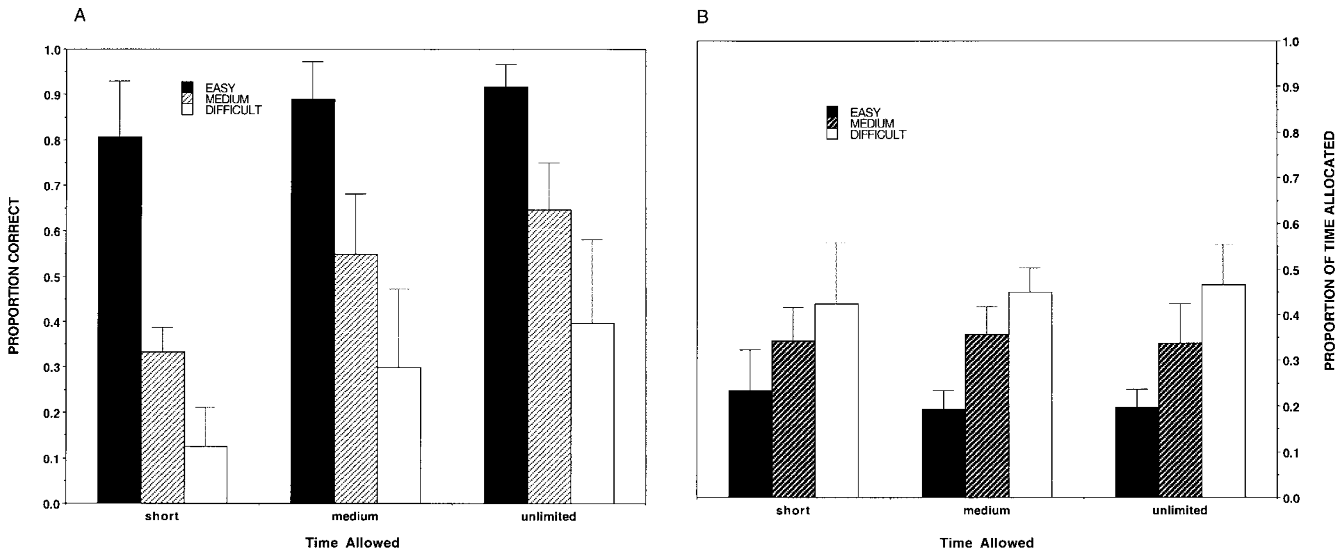

Figure 2. Experiment 2. A: Recall, scored leniently, of easy, medium, and difficult items. B: Proportionate time devoted to easy, medium, and difficult items for the true experts. Error bars indicate 95% confidence intervals.

the ceiling is that some of the words had multiple translations, and the computer scoring accepted only the translation that was designated. Thus, the levels of performance shown in this experiment—of about 90% in the easy conditions—are an underestimation.) When the experts were allowed more study time, their performance improved, $F(2, 16) = 10.81$, $MSE = .04$, ES = .57, for lenient scoring, and $F(2, 16) = 13.11$, $MSE = .02$, ES = .62, for strict scoring.

Study-Time Allocation

The study-time patterns in this experiment were different from those seen in Experiment 1. Experts devoted more of their study time to the medium and difficult items and less of their time to the easy items. The data on proportion of total time, given in Figure 2B, showed an effect of difficulty, $F(2, 16) = 11.91$, $MSE = .03$, ES = .59. Tukey tests showed that the difficult items were studied longer than the easy items, but there was no significant difference between time allocated to easy and medium items or to medium and difficult items. There was no interaction between time allowed and difficulty with the experts when the time data were broken out into the proportion of the total time spent that was devoted to each item type ($F < 1$). The experts, even with small amounts of time, concentrated on the difficult items, and no further relative shift toward the difficult items was evidenced when more time was allowed.

Comparison of Experts and Novices: Experiments 1 and 2

Recall performance. The difference in recall performance between the 9 expert subjects from Experiment 2 and the 12 novice subjects from Experiment 1 (the group factor) was significant, $F(1, 19) = 25.98$, $MSE = .10$, ES = .58, for lenient scoring, and $F(1, 19) = 36.98$, $MSE = .10$, ES = .66, for strict scoring, with the experts performing better than the novices. With strict scoring, there was an interaction between difficulty and group, $F(2, 38) = 3.93$, $MSE = .03$, ES = .17, and also an interaction between difficulty, time allowed, and group, $F(4, 76) = 4.22$, $MSE = .01$, ES = .18, suggesting that the novices were selectively impaired on the difficult items and that the allowance of longer time did not help the novices on these items, though it did help the experts' performance. Neither of these interactions was significant with lenient scoring, however, and so they are not interpreted further.

Study-time allocation. There was a main effect of group when study time was analyzed unconditionally, such that the novices spent more time overall than did the experts, $F(1, 19) = 12.33$, $MSE = 13.99$, ES = .39. There was also an interaction between difficulty and group, such that the novices spent their time on easier items than did the experts, $F(2, 38) = 4.36$, $MSE = 7.18$, ES = .19. Although there was no between-groups difference in the amount of time spent on the easy items ($t < 1$), the novices spent less time on the difficult items than did the experts, $t(19) = -3.1$, $p = .01$. When the study-time data were analyzed on the basis of proportion of total time in each condition spent on each item type, the effect of time allowed, as well as its interaction with group, was conditionalized away. However, the crucial interaction between difficulty and group was significant, $F(2, 38) = 10.24$, $MSE = .04$, ES = .35. The novices devoted proportionately more of their time to easier items than did the experts, $t(19) = -2.33$, $p = .03$; whereas experts spent proportionately more time on the difficult items than did novices, $t(19) = 3.78$, $p = .00$. This interaction is the expert shift predicted by the region of proximal learning hypothesis. The effects of difficulty and the interaction between difficulty and time allowed were significant in this joint analysis, but no other interactions with group were reliable. The data on the discarded participants are shown in Appendix B.

Discussion

The results contrasting novice- with expert-study-time behavior suggested that when the knowledge base shifted, so, too, did the strategy. Experts tended to study those items that were in their own perceived region of proximal learning, which tended to be the difficult items. Novices spent their time on easier items. It is interesting to note that the perception of what items constitute one's region of proximal learning appears to sometimes be incorrect, as illustrated in the data of the pseudoexperts, who thought they were experts by their own self-report and allocated their time like real experts but whose recall performance was like that of novices.

Experiment 3

The expert shift findings from Experiments 1 and 2 suggest that those supposedly null results from studies with mildly retarded teenagers and young children might actually have indicated an adaptive strategy whereby the young participants were devoting their time to those items that they perceived were in their own regions of proximal learning. To further investigate the possibility that expertise differences might result in a leftward shift with young participants, in Experiment 3 the children were novices in Spanish. In Experiment 4 they were experts. The hypothesis was that the novice children would behave in a manner similar to that of the novice adults, devoting their time selectively to the easy items. The expert children, moreover, were expected to allocate their study time like the expert adults—to the more difficult items. Such a pattern of results would provide support for an alternative interpretation of the previous null results.

Method

The design and procedures in this experiment were identical to those used in Experiments 1 and 2. The only difference was that the participants were 10 non-Spanish-speaking Grade 6 children, from a public middle school in the New York City school system, who were treated in accordance with the ethical principles of the American Psychological Association.

Results

Recall Performance

The children's recall data, scored leniently, are shown in Figure 3A, and, the recall data scored strictly are shown in Table 1. There was an effect of difficulty, $F(2, 18) = 61.19$, $MSE = .01$, ES = .87, for lenient scoring, and $F(2, 18) = 41.16$, $MSE = .01$, ES = .82, for strict scoring. Tukey tests for both strict and lenient scoring showed that participants performed better on the easy than medium items and that they did better on the easy than difficult items. The difference between the medium and difficult items was not reliable. Performance improved as the children were given more time, $F(2, 18) = 8.11$, $MSE = .01$, ES = .47, for lenient scoring, and $F(2, 18) = 13.35$, $MSE = .00$, ES = .60, for strict scoring, and there was an interaction in performance improvement between the difficulty of the item and the time allowed, $F(4, 36) = 6.468$, $MSE = .01$, ES = .42, for lenient scoring, and $F(4, 36) = 9.74$, $MSE = .00$, ES = .52, for strict scoring. As the children were given more time, they were more able to learn the easy items, but even considerable amounts of additional time had little effect on the medium and, especially, the difficult items. The difference between 5 and 60 s was reliable for the easy items, $t(9) = -4.67$, $p = .00$, for lenient scoring, and $t(9) = -4.98$, $p = .00$, for strict scoring, whereas neither of these two contrasts between short and long study times approached significance with items of medium difficulty. The statistic could not be computed because of lack of nonzero data for the difficult items.

Study-Time Allocation

When the amount of time spent on each item type was computed as a proportion of the total time spent studying in each triad, as shown in Figure 3B, the effect of difficulty was significant, $F(2, 18) = 3.84$, $MSE = .04$, ES = .30, though none of the individual comparisons were significant by Tukey tests. The crucial interaction between time allowed and difficulty was also significant, *F*(4,

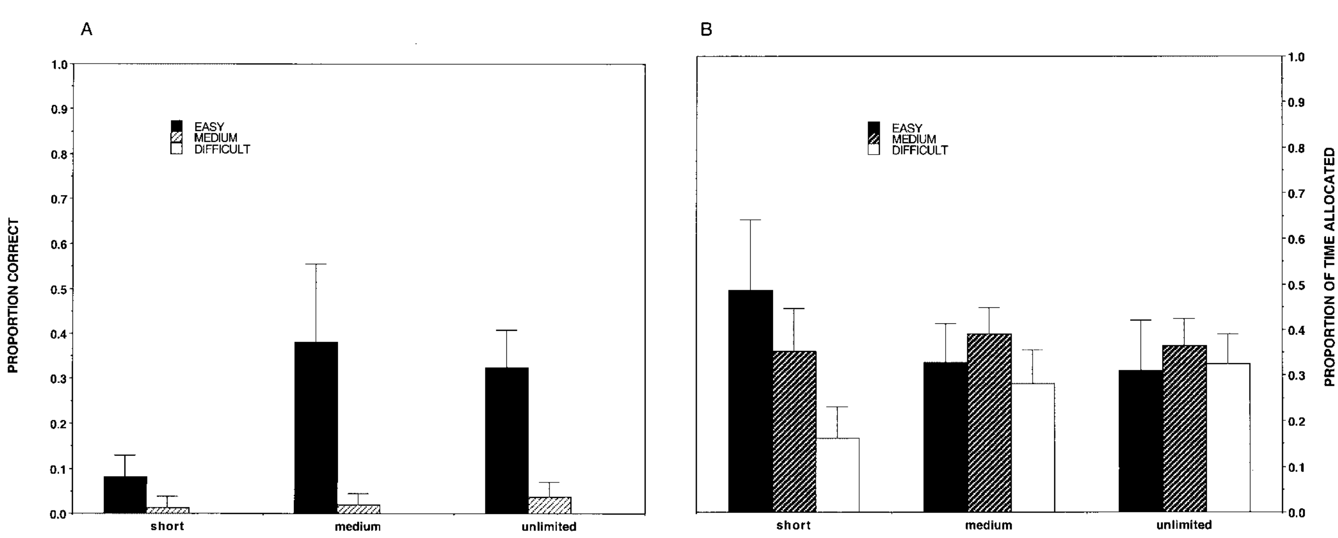

Figure 3. Experiment 3. A: Recall, scored leniently, of easy, medium, and difficult items. B: Proportionate time devoted to easy, medium, and difficult items. Error bars indicate 95% confidence intervals.

36) = 2.82, *MSE* = .03, ES = .24. Although the children favored the easy items over the difficult items at the 5-s rate, $t(9) = 2.92$, $p = .02$, they did not do so at either the 15-s or the unlimited rate, both *t*s were < 1.

Comparison of Experiments 1 and 3: Adult Novices Versus Child Novices

Recall performance. The child novices performed less well than did the adult novices, $F(1, 20) = 36.52$, *MSE* = .06, ES = .65, for lenient scoring, and $F(1, 20) = 19.73$, *MSE* = .05, ES = .50, for strict scoring, as would be expected on many grounds. There was also a significant interaction between difficulty and group, $F(2, 40) = 24.98$, *MSE* = .02, ES = .56, for lenient scoring, and $F(2, 40) = 16.74$, *MSE* = .02, ES = .46, for strict scoring. This interaction is likely due to floor effects on the difficult items, however, and so is not interpreted further. Although both age groups benefited from more time, there was a trend suggesting that the adults got more benefit from additional time with lenient scoring, which was not apparent when the data were scored strictly, $F(2, 40) = 2.93$, *MSE* = .01, ES = .13, $p = .07$, for lenient scoring. (As will be seen shortly, this may be because the adults spent more time studying in the unlimited-time condition, than did the children.) Similarly, the interaction among difficulty, time allowed, and group was significant when performance was scored leniently, $F(4, 80) = 2.54$, *MSE* = .01, ES = .11, but did not approach significance ($F < 1$) when scored strictly.

Study-time allocation. There was an effect of group, $F(1, 20) = 8.38$, *MSE* = 16.01, ES = .29, and an interaction between group and difficulty, $F(2, 40) = 3.18$, *MSE* = 6.63, ES = .14, when study time was assessed unconditionally. Although the difference between time devoted to the easy items did not differ between the two groups, $t(20) = .60$, $p = .55$, there was a difference (favoring the older participants) in time devoted to the difficult items, $t(20) = 2.39$, $p = .03$. There was also an interaction between time allowed and group, $F(2, 40) = 8.02$, *MSE* = 15.89, ES = .29. The adults tended to spend more time studying than the children, particularly in the unlimited-time condition: The *t* values for the short and medium condition were less than 1, whereas on the unlimited-time condition, $t(20) = 2.87$, $p = .01$, suggesting that the children may have given up or become impatient to a greater extent than did the adult novices. With the proportionate time analysis, no age-related interactions were significant. As would be expected from the results of the experiments taken separately, both groups showed the shift toward the more difficult items as more time was allowed. The interaction between time allowed and difficulty was significant, $F(4, 80) = 8.2$, *MSE* = .03, ES = .29.

Discussion

The results of this experiment, in conjunction with those of Experiment 1, are especially interesting, insofar as they showed that the children were behaving much like the adult novices. They did not selectively study the most difficult items (as the discrepancy reduction theory predicts for everyone) but rather showed apparently strategic behavior consistent with the region of proximal learning hypothesis, doing basically what the adults (who were also novices) did, under similar conditions. (Had we compared child novices, with adult experts, though, our results would have seemed consistent with the past literature.) Indeed, insofar as the children's performance in this experiment was poorer than that of the novice adults, the region of proximal learning theory would suggest that their region might even be toward easier items than that of the adult novices; and, indeed, when their raw study-time allocation data were compared to those of the adults, the adult novices did devote their time to harder items than did the children.

Experiment 4

It is tempting to interpret the young children's pattern of study-time allocation in a manner that is favorable to the region of proximal learning hypothesis—as indicating that the children's region of proximal learning was somewhat lower than that of the adults (as is consistent with their performance data) and that they had appropriately homed in on those items that they were most likely to learn. However, an alternative interpretation exists. Perhaps the children, being young and metacognitively unsophisticated, chose the wrong items to study because of some metacognitive deficit, resulting in poorer learning performance. Thus, although the interpretation suggested by the region of proximal learning hypothesis seems both plausible and appealing, it is not entirely unimpeachable.

This interpretation rests on the assumption that young children from this population will tend to choose the easy items or to be indiscriminate because of a failure in assessment or metacognition. If so, then one would expect that other children, from the same population, but who happened to be more expert in the domain, would also choose those items to study. If, however, more expert children showed a pattern reflecting their expertise, it would lend credence to the idea that the study-time pattern shown by the non-Spanish children was strategic. Thus, in Experiment 4, children from the same population and school, but who were experts in Spanish, were tested.

Method

The method in this experiment was identical to that used in the previous ones except that the 11 children were fluent Spanish speakers. They were selected from the school's bilingual class, which includes students whose first language is Spanish and who are in the process of learning English. Most of these children were able to understand English fairly well, though in some cases their lack of facility with what is here considered the base language, English, may account for some apparent learning difficulties.

Results

Recall Performance

There was an effect of difficulty on the expert children's recall performance, both when scored leniently, $F(2, 20) = 129.63$, *MSE* = .02, ES = .93, and when scored strictly, $F(2, 20) = 97.35$, *MSE* = .02, ES = .91. Tukey tests showed that they performed better on the easy than the medium items, better on the easy than the difficult items, and better on the medium than the difficult items, for both lenient and strict scoring. No other performance effects were significant. The performance data for strict scoring are given in Table 1. The leniently scored data are given in Figure 4B.

Study-Time Allocation

As can be seen from the proportionate analyses shown in Figure 4B, the expert children devoted time selectively to the difficult

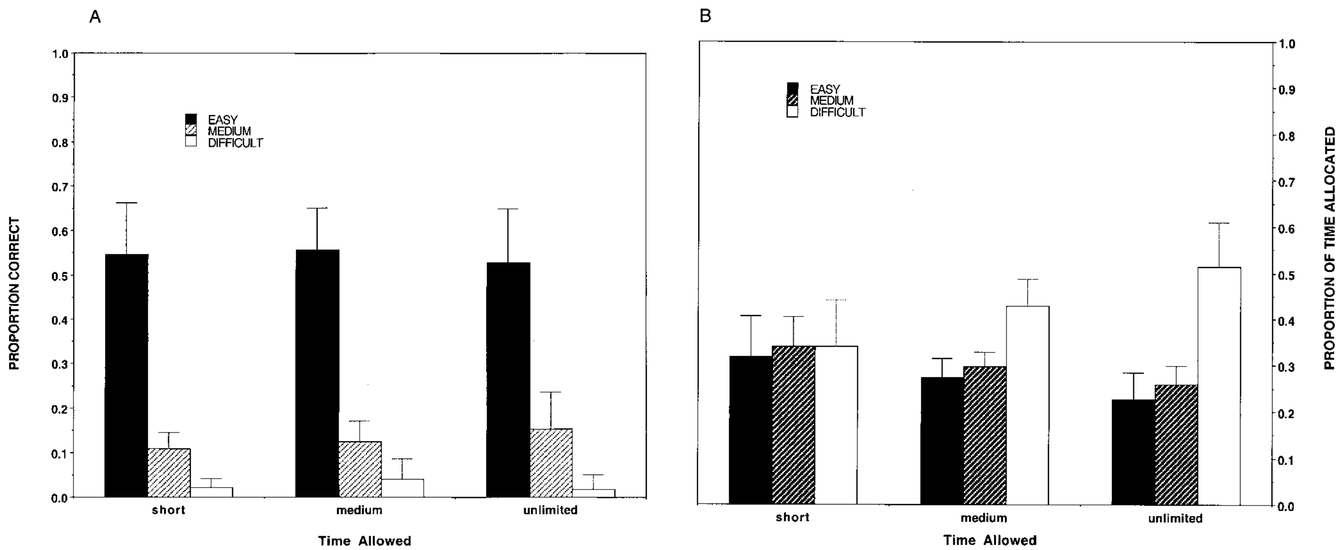

Figure 4. Experiment 4. A: Recall, scored leniently, of easy, medium, and difficult items. B: Proportionate time devoted to easy, medium, and difficult items. Error bars indicate 95% confidence intervals.

items, $F(2, 20) = 5.54$, $MSE = .04$, ES= .36, with post hoc analyses showing that they devoted more time to difficult than to easy items and more time to difficult than to medium-difficulty items, but there was not a significant difference between the easy items and the items of medium difficulty. There was also an interaction between difficulty and time allowed, $F(4, 40) = 6.03$, $MSE = .01$, ES = .38. Post hoc t tests on the interaction showed that there was no difference in time allocated to the easy and difficult items in the 5-s condition , $t < 1$, whereas in both the 15-s and the unlimited-time condition, more time was given to the difficult items, $t(10) = -3.14$, $p = .01$, and $t(10) = -3.67$, $p = .00$, respectively.

Comparison of Experiments 3 and 4: Child Experts and Novices

Recall performance. A comparison between the recall data from Experiments 3 and 4 revealed an effect of group, showing that the expert children performed better than did the novices, both when the data were analyzed leniently, $F(1, 19) = 22.46$, $MSE = .04$, ES = .54, and strictly, $F(1, 19) = 29.30$, $MSE = .03$, ES = .61. There was a difficulty by group interaction, with both lenient, $F(2, 38) = 17.83$, $MSE = .02$, ES = .48, and strict scoring, $F(2, 38) = 28.07$, $MSE = .01$, ES = .60. Tukey tests showed that the experts performed better than the novices at all levels of difficulty when leniently scored and better than the novices on the easy and medium difficulty items when strictly scored. The leniently-scored data showed an interaction between time allowed and group, $F(2, 38) = 5.2$, $MSE = .04$, ES = .14, suggesting that the additional time helped the novices more than the experts. This effect was not significant with the strict scoring, however, and is not interpreted further here. With both lenient and strict scoring, there was a significant triple interaction between difficulty, time allowed, and group, $F(4, 76) = 5.41$, $MSE = .01$, ES = .22, for lenient scoring, $F(4, 76) = 4.62$, $MSE = .01$, ES = .20, for strict scoring. As can be seen from Table 1, and from Figures 3B and 4B, extra time allowed (operationalized as the difference in performance between the 5-s and the unlimited-time conditions) did not benefit the expert children on the easy items (ts < 1 for both strict and lenient scoring), whereas it produced a benefit for the novice children on those same easy items, $t(19) = 5.16$, $p = .00$, for lenient scoring, and $t(19) = 5.34$, $p = .00$, for strict scoring. There are two reasons for this triple interaction. The first is that the expert children knew or thought they knew the easy items already. The second is that the expert children devoted less time to the easy items than did the novice children, presumably because they thought (sometimes wrongly) that they knew them already. The novice children devoted their extra time to the easy items and benefited by performance gains on those items.

Study-time allocation. The patterns of study time, as comparison of Figures 3B and 4B reveal, were different for the two groups. Although there was no difference between the two groups in the total amount of time devoted overall, in the unconditional analysis, the crucial interaction between difficulty and group was significant, $F(2, 38) = 9.17$, $MSE = 2.45$, ES = .33. Post hoc t tests showed that although there was no difference between the two groups on the easy items, there was a difference on the medium items, $t(19) = -2.54$, $p = .02$, and although the difference was not significant between the two groups on the difficult items, it was in the opposite direction, $t(19) = 1.37$, $p = .18$. This interaction—the critical expert shift—was also significant when the data were analyzed in terms of proportionate time, $F(2, 38) = 9.22$, $MSE = .04$, ES = .33. This time, all three post hoc t tests were significant, with the direction on the easy and medium items favoring the novices, $t(19) = -2.32$, $p = .03$, and $t(19) = -2.72$, $p = .01$, respectively, whereas the comparison on the difficult items favored the experts, $t(19) = 3.62$, $p = .00$. No other interaction with expertise group on study-time allocation was significant.

Discussion

The results of Experiment 4 indicated that the children, like the adults, were behaving strategically, even though, in the case of the expert children, the strategies were not always very effective. The children in Experiment 3 presumably were not simply using an inappropriate or random strategy because their metacognitive or control processes were lacking. If this were the reason for their selection of items, then one would expect a similar response from children from the same population who happened to be experts. However, the children from this pool who were experts in the particular domain showed the expert shift, just as did the adult experts as compared with the adult novices. It is notable, however, that the expert shift may have led the expert children to labor in vain, insofar as their recall data indicate that there was considerable room for improvement on the items of easy and medium difficulty. Their emphasis on the most difficult items, especially at the slow-presentation rate, produced almost no return in their later recall performance. So it would appear that although they may have been trying to isolate their own region of proximal learning, they were not doing so with great success.

One final analysis was conducted by combining proportionate study-time data on all four experiments in a 2 (age: either college students or children) × 2 (expertise: either novices or experts) × 3 (difficulty: easy, medium, or difficult) × 3 (time allowed : 5 s,15 s, or unlimited) split-plot design. This combined analysis allows investigation of the major effects across the entire sequence of experiments. The results were remarkably simple. First, there was the interaction between expertise and difficulty, $F(2, 76) = 19.38$, $MSE = .04$, ES = .34. The experts studied more difficult items than did the novices. Second, the interaction between age and difficulty was also significant, $F(2, 76) = 3.53$, $MSE = .04$, ES = .08. The children tended to allocate their time to easier items than did the college students. The age difference is probably not simply due to age, per se, but rather to the fact that the children, overall, had a perceived region of proximal learning that was skewed toward the easier items, presumably because their learning was less advanced than the college students (as their performance also showed). Both of these interactions qualified a simple main effect of difficulty, $F(2, 76) = 5.17$, $MSE = .04$, ES = .12. Finally, the interaction between difficulty and time allowed was reliable, $F(4, 152) = 10.49$, $MSE = .02$, ES = .22. As people had more time to study, their allocation of study time shifted toward the more difficult items. No other effects or interactions were significant.

Experiment 5

In this experiment, one final prediction was tested: Over trials the region in which time is selectively allocated should shift toward the more difficult items. The rationale for this prediction was identical to that involving study time—that is, as learning advances, in this case with each additional study trial, more and more of the easy items should enter the learned state, and, therefore, the region in which it is advantageous to devote learning effort and study time should shift to those more difficult items that have not yet been learned.

Method

The method in this experiment was identical to that in the four previous experiments, with the following modifications. First, only one time-allowed condition—the 5-s condition—was used. This allowed completion of the task within 1 hr. Second, participants were given four study-test trials, rather than one. They first studied the items by clicking on the translations they wished to see, and, then, at the end of the study phase, they were tested in a random order on all 144 items. Then the computer rerandomized the orders and presented three additional study-test trials on the same materials. The participants were 12 Columbia University students who received a small bonus course credit for participating.

Results

Recall Performance

As can be seen from Table 2 and Figure 5A, recall performance was best for the easy items, next best for the items of medium difficulty, and worst for the difficult items, when assessed both leniently, $F(2, 22) = 197.97$, $MSE = .03$, ES = .95, and strictly, $F(2, 22) = 128.34$, $MSE = .02$, ES = .92. Tukey tests with both strict and lenient scoring showed that participants did better on easy than on medium-difficulty items and better on medium than on difficult items. There was also an effect of trial, $F(3, 33) = 147.13$, $MSE = .00$, ES = .93, for lenient scoring, and $F(3, 33) = 99.52$, $MSE = .00$, ES = .90, for strict scoring. Tukey tests showed that for both strict and lenient scoring, people did better on Trial 2 than on Trial 1, better on Trial 3 than on Trial 2, and better on Trial 4 than on Trial 3 (as well as on all pairs that were more than one step removed, of course). Finally, there was an interaction between difficulty and trial, $F(6, 66) = 35.98$, $MSE = .00$, ES = .77, for lenient scoring, and $F(6, 66) = 33.71$, $MSE = .00$, ES = .75, for strict scoring. There was almost no change in accuracy on the difficult items, $t(11) = -1.60$, for lenient scoring, and $t(11) = 1.48$, for strict comparison between Trial 1 and 4, both nonsignificant, whereas there was considerable improvement for both the easy and medium difficulty items over trials, $t(11) = -9.76$, $p = .00$, for lenient scoring, $t(11) = -11.41$, $p = .00$, for strict scoring, comparing Trials 1 and 4 on the easy items, and $t(11) = -9.14$, $p = .00$, for lenient scoring, and $t(11) = -6.58$, $p = .00$, for strict scoring, comparing Trial 1 and Trial 4 on the medium items.

Study-Time Allocation

Figure 5B shows the proportionate times on each item type over trials (the absolute data is not presented here because the total time

Table 2
Strictly Scored Recall Performance for Experiment 5 (and 95% Confidence Intervals)

Item difficulty	Trial 1	Trial 2	Trial 3	Trial 4
Easy	0.288 (0.082)	0.474 (0.083)	0.597 (0.087)	0.656 (0.079)
Medium	0.050 (0.025)	0.151 (0.042)	0.259 (0.052)	0.313 (0.084)
Difficult	0.000 (0.000)	0.002 (0.003)	0.005 (0.005)	0.007 (0.009)

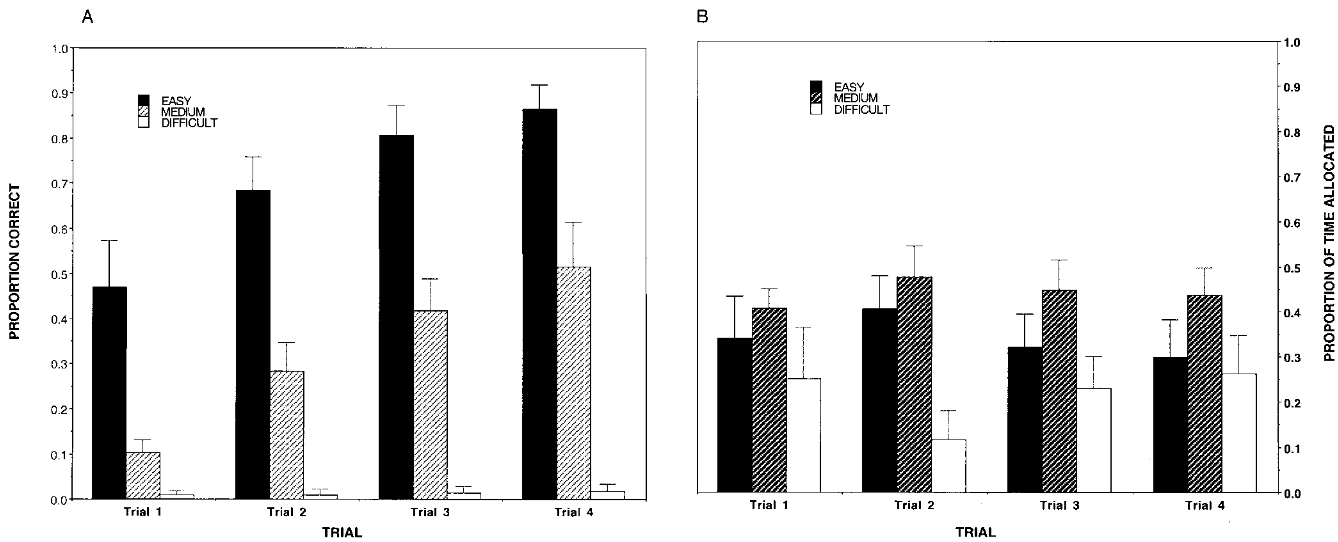

Figure 5. Experiment 5. A: Recall, scored leniently, of easy, medium, and difficult items. B: Proportionate time devoted to easy, medium, and difficult items. Error bars indicate 95% confidence intervals.

was restricted to 5 s, and so the two patterns of data are similar). On the first trial, the data are quite consistent with the 5-s condition in Experiment 1, as would be expected. The participants allocated their time to the easy and medium difficulty items, predominantly. However, in contrast to the predictions given above, study time on the second trial did not shift toward the more difficult items. Instead, on Trial 2, there was a regressive shift, on average, toward studying items that were even easier than had been the case on Trial 1. Trials 3 and 4 showed a pattern that was about the same as the original pattern on Trial 1, but which, in the context of the regressive shift on Trial 2, appears to be the start of a shift toward the more difficult items. Statistically, there was an effect of difficulty, showing that more time was allocated to the easy and moderate difficulty items, overall, $F(2, 22) = 10.59$, $MSE = .59$, ES $= .50$. There was, of course, no effect of trial. There was, however, a significant interaction between difficulty and trial, $F(6, 66) = 2.60$, $MSE = .02$, ES $= .19$. To further analyze the locus of this interaction, a summary statistic was calculated: time allocated $(\text{difficult}-\text{easy})_{\text{on Trial } n}$ − time allocated $(\text{difficult}-\text{easy})_{\text{on Trial } n-1}$. A positive value indicates that time was increasingly being allocated to more difficult items in the later trial. A zero value indicates no change in time allocation. A negative value indicates a regressive shift, whereby time was being allocated to easier items on the later trial. Three contrasts were computed—between Trials 1 and 2, between Trials 2 and 3, and between Trials 3 and 4—and post hoc tests compared to zero were computed. The value of the Trial 1 to 2 contrast was significantly negative, $t(11) = -2.43$, $p = .03$, indicating a reliable regressive shift. The contrast between Trials 2 and 3 was significantly positive, $t(11) = 2.53$, $p = .03$, indicating a difficulty shift; and the contrast between Trial 3 and 4 was also positive but not significantly so, $t(11) = 1.14$, $p = .28$.

Discussion

In this experiment, the straightforward prediction of the region of proximal learning hypothesis—that over trials there would be a shift in study-time allocation to the right—was not borne out, but perhaps with good reason. This was the only experiment in the series in which people received any feedback about their learning performance before they had a chance to restudy. Although they were not explicitly given correction, when they took the test at the end of each trial, participants presumably had the opportunity to experience their own degree of learning and to realize that very frequently, they could not come up with easy items. Learning, at the end of Trial 1, even on the easiest items, was not impressive: Scored strictly, it was only 29% (or nearly 50% when scored leniently, which may be closer to the participants' perceived performance). Being able to remember only 50% of the items in a difficulty class, though, is far from the complete learning that—according to the region of proximal learning hypothesis—would result in a shift in emphasis toward more difficult items. It seems likely that once the participants realized that they had not yet mastered the items of easy and moderate difficulty and that there was plenty of room for improvement in this region of relatively easy pickings, rather than shifting their time to yet harder items which would offer even less opportunity for overall performance gains, they opted to devote even more time to the easy and moderate as-yet-unlearned translations.

The finding of a regressive shift in study-time allocation is consistent with recent findings reported by Koriat, Sheffer, and Ma'ayan (2002), who showed that there is systematic underconfidence in people's ratings on Trial 2. Although it is well established that there is an overconfidence effect on Trial 1, this new finding accords well with the present result of a concomitant regression toward studying easier items. As more of the easy items entered the learned state at the end of Trial 2, however (as evidenced by performance on those items of over 70%, as shown in Figure 5), people started to give relatively less time to them and relatively more time to the difficult items. Performance on the items of moderate difficulty never approached ceiling, and people continued to devote more of their study time to those items than to other items, even on the fourth learning trial. Consistent with

Nelson and Leonesio's (l988) early findings on the labor-in-vain effect, considerable quantities of additional time on the difficult items on Trials 3 and 4 yielded unimpressive performance gains on those items. Thus, although the overall pattern of results was inconsistent with the simple a priori predictions that were based on the assumption of very high levels of learning of the easy items following Trial 1 (an assumption that did not materialize in the data), it is completely reasonable in light of the performance data that ensued. The predictions, then, failed to materialize, not because people were not attempting to home in on a region of proximal learning but rather because they were.

Conclusion

The results from these five experiments favor the region of proximal learning hypothesis. Consistent with that hypothesis, people tried to selectively allocate their time to the items that would yield the maximum learning return for them. Unless they were adult experts, people did not allocate most of their study time to the most difficult items. Rather, given little study time they devoted that time to the items that were of easy and medium difficulty, presumably those that they believed would be most likely to be learned with some effort. When they had more time to study, people's study-time allocation shifted to items that were of greater difficulty, presumably because some of the easy items had either already been learned or because the participants (sometimes mistakenly) believed that they had been learned, leaving time to devote to items that were more difficult. When people were true experts, rather than novices, they devoted their time to the more difficult items—the only items, that for them, were still to be learned. Pseudoexperts behaved the same way, but for them, the strategy appeared to mismatch their knowledge. When people were given multiple study trials and discovered after the first trial that their performance on the easy and medium-difficulty items was not near ceiling, they shifted their study-time allocation strategies downward to devote time to learning those relatively easy items that were still to be learned. Once performance improved, they then started to shift toward the more difficult items, as the theory would predict. This entire pattern echoes the learning dynamic expected by the region of proximal learning hypothesis.

Several of these results are at odds with the discrepancy reduction model, which was previously thought to explain virtually all of the findings in the field. First, the finding that when time is limited, people devote their time selectively to the items of easy and medium difficulty rather than to the most difficult items conflicts with the basic tenet of the discrepancy reduction model. Second, the finding that children who chose to study the easier items were probably behaving strategically (since expert children do otherwise) is both inconsistent with the discrepancy reduction model and suggests that some of the exceptions in the previous literature were not necessarily failures of metacognition or control. Rather the current finding, and these past exceptions, point to the possibility of a fundamentally different underlying mechanism than that offered by the discrepancy reduction model. Third, the discrepancy reduction model offers no principled reason why the amount of study time allowed should systematically affect which items the person chooses to study. All of these findings, are, however, consistent with the idea that people are according study time within their own perceived region of proximal learning, a region that changes with their expertise and with the time allowed.

The region of proximal learning hypothesis is not embarrassed by the findings that under some circumstances, people will allocate their time selectively to the most difficult items—the basic data that originally led to the discrepancy reduction model but which must also be accounted for by any explanation of study-time allocation. Under conditions in which the most difficult items happen to be those items in an individual's region of proximal learning, of course, one would expect people to devote their time to them. Those specific conditions are (a) when the person is an expert and already has learned the items of easy and medium difficulty and (b) when the person is given unlimited time (or unlimited study opportunities) and hence has learned the easier items. Thus, the pattern of results predicted by the discrepancy reduction model is a special case.

Many learning theorists have postulated that people progress, cognitively, by directing their efforts to learning those materials that are just beyond what they have currently mastered. Learning is seen as a dynamic bootstrapping operation, with each step in the process dependent upon the last. The past literature on study-time allocation seemed inconsistent with this presumed fundamental mode of learning. The conflict between what people should study and what they apparently did study seemed to justify concerns that people's metacognitions were systematically misleading and untrustworthy; the metacognitive and control strategies observed in the past study-time allocation literature seemed dysfunctional. The results presented in this article, in contrast, indicate that people do try to isolate their own region of proximal learning and to study selectively within it.

References

Atkinson, R. C. (l972a). Ingredients for a theory of instruction. *American Psychologist,* 921–931.

Atkinson, R. C. (l972b). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96,* 124–129.

Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology, 2,* 411–420.

Berlyne, D. E. (l978). Curiosity and learning. *Motivation and Emotion, 2,* 97–175.

Bisanz, G. L., Vesonder, G. T., & Voss, J. F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning: A developmental study. *Journal of Experimental Child Psychology, 25,* 116–128.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition, 22,* 249–257.

Dufresne, A., & Kobasigawa, A. (1988). Developmental differences in children's spontaneous allocation of study time. *The Journal of Genetic Psychology, 149,* 87–92.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 47,* 274–296.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25,* 691–700.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journal of Gerontology: Psychological Science, 52,* 178–186.

Dunlosky, J., & Hertzog, C. (l998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.

Hebb, D. O. (l949). *Organization of behavior.* New York: Wiley.

Kellas, G., & Butterfield, E. C. (1971). Effect of response requirement and type of material on acquisition and retention performance in short-term memory. *Journal of Experimental Psychology, 88,* 50–56.

Kobasigawa, A., & Dufresne, A. (1992). Differential allocation of study time by Grade 3 children. Unpublished manuscript.

Kobasigawa, A., & Metcalf-Haggert, A. (1993). Spontaneous allocation of study time by first- and third-grade children in a simple memory task. *The Journal of Genetic Psychology, 154,* 223–235.

Koriat, A. (l993). How do we know what we know? The accessibility model of feeling of knowing. *Psychological Review, 100,* 609–639.

Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115–135). Cambridge, MA: MIT Press.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9,* 149–171.

Koriat, A., & Goldsmith, M. (l996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131,* 147–162.

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica, 36,* 280–289.

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15,* 237–246.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study-time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122,* 47–60.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18,* 196–204.

Mazzoni, G., Cornoldi, C., Tomat, L., & Vecchi, T. (1997). Remembering the grocery shopping list: A study on metacognitive biases. *Applied Cognitive Psychology, 11,* 253–267.

Metcalfe, J. (l993a). Monitoring and gain control in an episodic memory model: Relation to P300 event-related potentials. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 327–354). Hillsdale, NJ: Erlbaum.

Metcalfe, J. (1993b). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100,* 3–22.

Metcalfe, J. (l996). Metacognition. In E. L. Bjork & R. A. Bjork (Eds.), *The handbook of perception and cognition: Vol. 10. Memory* (pp. 383–411). San Diego, CA: Academic Press.

Metcalfe, J. (l998). Insight and metacognition. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology* (pp. 181–197). Mahwah, NJ: Erlbaum.

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 197–211). Cambridge, UK: Oxford University Press.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (l993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851–861.

Metcalfe, J., & Shimamura, A. S. (l994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

Miner, A. C., & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47–70). Cambridge, MA: MIT Press.

Nelson, T. O. (l990). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General, 112,* 269–273.

Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3,* 317–318.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207–213.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686.

Nelson, T. O., & Narens, L. (l994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–26). Cambridge, MA: MIT Press.

Piaget, J. (l952) *The origins of intelligence in children.* New York: International Universities Press.

Reder, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 227–259). San Diego, CA: Academic Press.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435–452.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 18,* 1074–1083.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–114). Cambridge, MA: MIT Press.

Schwartz, B. L., & Smith, S. (l997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory and Language, 36,* 68–86.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221.

Thiede, K. W. (l996). The relative importance of anticipated test format and anticipated difficulty on performance. *The Quarterly Journal of Experimental Psychology, 49A,* 901–918.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037.

Vygotsky, L. S. (l987). *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology including the volume thinking and speech.* New York: Plenum Press.

Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology, 82,* 441–447.

(*Appendixes follow*)

Appendix A

Easy, Medium, and Difficult Spanish–English Translations

Spanish	English
Easy pairs	
fantastico	fantastic
intensivo	intensive
intervalo	interval
familia	family
aspecto	aspect
asteroide	asteroid
aspirina	aspirin
bateria	battery
biografia	biography
botella	bottle
boton	button
bufalo	buffalo
cacto	cactus
calcio	calcium
capitan	captain
defensivo	defensive
deduccion	deduction
grafica	graphic
deposito	deposit
depender	depend
domestico	domestic
ecologia	ecology
educacion	education
fraude	fraud
grasa	grease
enorme	enormous
entero	entire
fosil	fossil
fotografia	photograph
frase	phrase
fugitivo	fugitive
galaxia	galaxy
gargola	gargoyle
garantia	guarantee
gasolina	gasoline
generoso	generous
futuro	future
enciclopedia	encyclopedia
tomate	tomato
depresion	depression
enemigo	enemy
lampara	lamp
fortuna	fortune
aeropuerto	airport
leon	lion
majestad	majesty
ligamento	ligament
murmullo	murmur
Medium pairs	
infrahumano	subhuman
ingravidez	weightlessness
librecambio	free trade
marido	husband
noctambulo	sleepwalking
paloma	dove
picaporte	doorlatch
portaplumas	penholder
pluma	pen
rubor	blush
sepultura	burial
tapizado	carpeting
temprano	early
tentacion	temptation
ultimamente	lately
vagabundear	to wander
valentonada	boast
vaticinio	prophecy
viento	wind
vodevil	music hall
volver	turn
alimento	food
antiguo	ancient
aportar	furnish
aprovechar	exploit
brillantez	brightness
cargazon	shipment
carecer	to lack
cerveza	beer
centella	spark
cavernicola	caveman
cilindrin	cigarette
cimentacion	foundation
abrochar	to tie
desgracia	tragedy
desairado	unsuccessful
emisora	radio station
rabia	rage
enrabiar	enrage
fraccionamiento	division
empalmar	to join
embarazo	pregnancy
gomero	rubber tree
sangre	blood
terremoto	earthquake
labrador	farmer
lancha	motorboat
lavazas	dishwater
Difficult pairs	
chafarrinada	stain
guardaropa	closet
alcantarilla	drain
escaparate	window-shopping
jorguineria	sorcery
arandano	cranberry
abastecimiento	provision
boquiabierto	gaping
envenenamiento	poisoning
garrapaticida	insecticide
jerigonza	gibberish
lambisconeria	gluttony
mamposteria	masonry
otorgamiento	granting
prorrumpir	break out
sacapuntas	sharpener
socarroneria	sarcasm
trituradora	shredder
zanguango	slacker
anquilasamiento	stagnation
barrabasada	mischief
cerrajero	locksmith
desapercibido	unnoticed
embobamiento	fascination

Appendix A (*continued*)

Easy, Medium, and Difficult Spanish–English Translations

Word		Word	
Spanish	English	Spanish	English
Difficult pairs (*continued*)		Difficult pairs (*continued*)	
ferrocarril	railway	paliacate	handkerchief
gurrumino	weak	papanateria	gullibility
herramienta	tool	resquebrajadura	crack (n.)
jacarandoso	merry	tartamudez	stutter
marimorena	fuss	tejemaneje	bustle
cerbatana	blowpipe	buhardilla	skylight
cosechadora	harvester	aljofifa	floorcloth
herrero	blacksmith	choquezuela	kneecap
golondrina	swallow (n.)	descuajaringarse	crumble
gualdrapear	to flap	arrancaclavos	clawhammer
izquierdista	leftist	zangarriana	migraine
lengueterias	gossip	guadanadora	mower

Appendix B

The Participants Eliminated From Experiment 3

Eight participants were eliminated from the basic analysis of Experiment 2 because, although they were self-designated expert Spanish speakers, they had scored less than 30% correct on the strict accuracy performance test. This group is potentially interesting insofar as one possible factor contributing to its poor performance may have been an impairment in metacognition and control strategies. These pseudoexperts were compared, by separate analyses of variance, to the novices from Exeriment 1 and to the true experts in Experiment 2.

The recall performance of the pseudoexperts was not significantly different from that of the novices from Experiment 1 (Fs < 1, both for strict and lenient accuracy). There was only one significant interaction between the novice and pseudoexpert groups with the accuracy measures and that was a triple interaction with strict scoring, among group, difficulty, and time allowed, $F(4, 72) = 2.92$, $MSE = .01$, ES = .14, but this interaction was not significant when the data were scored leniently. Basically, then, the novices' and pseudoexperts' recall performance was similar.

The recall performance of the pseudoexperts was worse than was that of the true experts, $F(1, 15) = 18.64$, $MSE = .11$, ES = .55, for lenient scoring, and $F(1, 15) = 25.48$, $MSE = .09$, ES = .63, for strict scoring. There was also an interaction between time allowed and group, which, although only a trend ($p = .10$) with leniently scored data, was significant with strict scoring, $F(2, 30) = 4.22$, $MSE = .01$, ES = .22. The experts benefited more from the additional time allowed than did the pseudoexperts, who learned relatively little with additional time.

The novices devoted more time overall than did the pseudo experts, which showed up both as a main effect and as an interaction between group and time allowed with unconditional study-time allocation data. The novices devoted 11.27 s per item in the unlimited-time condition, whereas the pseudoexperts only devoted about 4.61 s, $F(2, 36) = 12.24$, $MSE = 17.34$, ES = .40. In the proportionate time analysis, there was an interaction between difficulty and group (as was predicted, if the pseudoexperts were behaving like the experts), $F(2, 36) = 2.83$, $MSE = .04$, ES = .14. There was no difference between the pseudoexperts and the novices on the easy items ($t < 1$), but the pseudoexperts devoted more time to the difficult items than did the novices, $t(18) = 1.9$, $p = .04$, one-tailed. The triple interaction between difficulty, time allowed, and group was marginal, $F(4, 72) = 2.30$, $MSE = .03$, $p = .07$, two-tailed, ES = .11. When the pseudoexperts were contrasted with the real experts, none of the effects or interactions approached significance, either with the unconditional time analysis or with the proportionate time analysis.

In summary, the pseudoexperts recall was like the novices, whereas their study-time allocation was like the experts. This pattern of poor performance coupled with an overestimation of their abilities and the same strategies as real experts suggests a metacognitive and control dysfunction. It would be of considerable interest to see if titrating the study-time allocation down to the group's real level of accuracy would result in improved performance.

Received June 31, 2001
Revision received February 28, 2002
Accepted February 28, 2002 ■