APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **19**: 107–122 (2005)
Published online 15 November 2004 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1066

Distributed and Massed Practice: From Laboratory to Classroom

RACHEL SEABROOK[1]*, GORDON D. A. BROWN[2] and JONATHAN E. SOLITY[2]

[1]*Oxford Brookes University, UK*
[2]*Warwick University, UK*

SUMMARY

The benefit to memory of spacing presentations of material is well established but lacks an adequate explanation and is rarely applied in education. This paper presents three experiments that examined the spacing effect and its application to education. Experiment 1 demonstrated that spacing repeated presentations of items is equally beneficial to memory for a wide range of ages, contrary to some theories. Experiment 2 introduced 'clustered' presentations as a more relevant control than massed, reflecting the fact that massed presentation of material is uncommon in education. The scheduling of clustered presentations was intermediate between massed and distributed, yet recall was no different than for massed. Experiment 3, a classroom-based study, demonstrated the benefit of distributed over clustered teaching of reading through modification of the scheduling of everyday lessons. Thus, the effectiveness of teaching may be improved by increasing the degree to which lessons are distributed.

The benefit for memory of distributed (over massed) presentation of to-be-remembered material is robust. Spaced presentation of material leads to better retentions than does blocked presentation. However the principle of distributing learning has not been widely applied in education. Dempster (1988) argued that a major factor limiting such application is the lack of direct classroom-relevant demonstrations of its efficacy. Indeed, it is questionable whether or not the contrast between distributed and massed presentation of material is relevant to education at all. Massed presentation, in the sense of repeated presentations of to-be-learnt stimuli, or more generally, concentrated teaching of a topic in a single session, is rarely found in the classroom. Of more interest is the question of whether different degrees of distribution are influential in learning. The experiments reported in this paper begin to address the need to bridge laboratory studies (Experiments 1 and 2) and classroom-based studies (Experiment 3). At the same time a more important comparison than between massed and distributed presentation is investigated, namely between different degrees of distribution.

*Correspondence to: Dr R. Seabrook, Department of Psychology, Oxford Brookes University, Gipsy Lane Campus, Oxford, OX3 0BP, UK. E-mail: rseabrook@brookes.ac.uk

Contract/grant sponsor: The Leverhulme Trust; contract/grant number: F/215/AY.

It is well established in research on adult memory that if a to-be-remembered list of words includes repeated items, not only will the repeated items be recalled (or recognized) better than the once-presented items, but also repetitions that are separated by other words will be better remembered than repetitions that occur in adjacent list positions (e.g. Glenberg, 1979). This is known as the spacing effect, or the benefit of distributed over massed presentation. Furthermore, the more items there are between two presentations of a repeated item, the better memory for that item will be. This is known as the lag effect, or sometimes the Melton effect (after Melton, 1970).

The benefit of distributed presentation has been found with a range of materials (e.g. episodic memory and paired associate learning for words: Glenberg, 1979; vocabulary learning: Dempster, 1987; memory for the main points of a text: Reder & Anderson, 1982; learning new concepts in biology: Reynolds & Glaser, 1964; and associating names with faces: Landauer & Bjork, 1978). It has been found with both recognition and recall (e.g. Greene, 1989), in both implicit and explicit memory (e.g. Greene, 1990) and following both intentional and incidental learning (Challis, 1993).

Such a robust effect has obvious potential value in education, and a number of studies have found the spacing effect with children of different ages, from infants (Cornell, 1980) through preschool and young primary school children (Cahill & Toppino, 1993; Toppino, 1991, 1993) to older children and adults (Rea & Modigliani, 1987; Toppino & DeMesquita, 1984; Toppino, Kasserman, & Mracek, 1991). In addition to these developmental studies, there have been examinations of spacing in specifically educational settings, such as those of Reynolds and Glaser (1964), who taught biology to secondary-school pupils using programmed instruction (machine-presented information); Rea and Modigliani (1985), who taught spelling and multiplication to primary-school children; and Dempster (1996), who taught new vocabulary to undergraduates.

Whilst the above studies demonstrate that the spacing effect can be found at all ages and with educationally realistic materials, a potential limitation has been highlighted by Wilson (1976). This relates to the lag effect (i.e. the generalization of the spacing effect such that increasing the gap or intervening information between presentations of an item further improves memory for that item). Wilson found that for the youngest children in his study a short lag was highly beneficial relative to massed presentations, but that there was no further benefit of a longer lag. Older children, in contrast, obtained additional benefit from longer lags. On the basis of these results, Wilson proposed an explanation of the lag effect in terms of working memory (WM) capacity (see Experiment 1 of this paper for details). His results were replicated by Toppino and DeMesquita (1984), who proposed an alternative explanation in terms of the developing use of organizational strategies in intentional learning.

Regardless of the theoretical explanation for this ostensible developmental difference, the apparent absence of an advantage of longer lags for younger children has strong implications for the applicability of the spacing effect to education. More specifically, for teaching of young children the implication of the spacing effect would be nothing more than that immediate repetition of material is inimical to learning. If, on the other hand, longer gaps between presentations are as beneficial to children as they are to adults, wider implications follow. Therefore, it is important to examine the possible differences in the lag effect across a wide age range. This was the purpose of Experiment 1, which compared lags of 0 (massed), 1, 3 and 8 intervening items in a list of words and tested participants ranging from 5-year-olds to undergraduates.

To anticipate the structure of the remainder of the paper: Further experiments extend the study of lag effects to include a more environmentally valid control condition and a more realistic situation. Experiment 2 introduces an intermediate mode of presentation, neither massed nor distributed, referred to as 'clustered' presentations, which include pairs of immediate repetitions as well as spacing between the pairs. This is in order to study conditions that, arguably, approximate educational reality more closely. Typically, teaching of any topic or skill will not be massed, but to some extent distributed. Therefore, from the point of view of classroom application, the interesting question is the extent to which different degrees of distribution influence subsequent memory. 'Clustering' is a 'less distributed' presentation schedule against which to compare highly distributed presentations.

In Experiment 3, the same methodology was extended to a classroom study, in which teachers were asked to modify the scheduling of their usual teaching of reading skills. This final experiment assessed, in the most direct way, the value of distributing teaching sessions in an educational setting.

EXPERIMENT 1

Whilst several published studies have found a spacing effect in young children's memory (Rea & Modigliani, 1987; Toppino, 1991), there is some evidence that the lag effect develops with age. Wilson (1976) studied this effect in 8- and 12-year-old children as well as in college students, and used lags of 0, 2 and 8 intervening items between repetitions. He found that whilst the 12-year-olds showed the same lag effect as adults, in that longer lags between repetitions led to better recall performance, the younger children showed a different pattern. For the 8-year-olds, there was a sharp increase in performance between lag 0 (i.e. immediate repetition) and lag 2 items, but no further increase with a lag of eight items.

What could give rise to this developmental difference? One theory of the spacing effect is that widely separating two presentations of an item increases the chances that they are encoded differently. This in turn increases the number of potential retrieval cues available for that item (e.g. Glenberg, 1979). Wilson (1976) argued that the probability of a new encoding at the second presentation depends on whether or not the item is still available in WM from the first presentation. If it is, then the original encoding will be repeated; otherwise, additional processing will be needed to re-encode the item. Therefore the lag effect is critically dependent of WM capacity. Since this develops with age, Wilson's account would predict that younger children, who have a lower capacity, would forget the first presentations more quickly and hence benefit from shorter lags than older children do. On the other hand, they would not benefit further from increasing lags.

One study, conducted by Toppino and DeMesquita (1984), has followed up Wilson's (1976) experiment. Toppino and DeMesquita studied 54 children aged 5-, 7- and 9-years old (18 in each year group)—a slightly younger range than in Wilson's experiment. Otherwise, the format of their experiment was essentially the same as Wilson's, with a slightly narrower range of lags: 0, 3 and 6 intervening items. They found that all the children benefited significantly from spaced presentations, with memory for lag 0 items being significantly worse than memory for items at either of the longer lags. In addition, they found no significant difference (for any of the age groups) between lag 3 and lag 6 items. This is the same pattern as Wilson found in the youngest group of children and, as

Toppino and DeMesquita argue, is different from what would be expected in adults. Thus the results of Toppino and DeMesquita appear to replicate Wilson's.

However, Toppino and DeMesquita (1984) did not find any differences between the age groups they studied. It is possible that their design did not contain enough detail to detect real differences and that significant effects would have been evident if shorter lags had been used. Furthermore, examination of their results reveals that there were differences (albeit non-significant) between lag 3 and lag 6 for all three age groups. These considerations highlight the need for a replication and extension of this work, which is the purpose of the current experiment.

Following the same paradigm as that used by Wilson (1976) and by Toppino and DeMesquita (1984), the current experiment allows a more fine-grained investigation of the differences between short lags by including lags of 0, 1, 3 and 8 intervening items. Secondly, a wider age range was covered, from 5-years-old to adult. This enables any differences to be detected between the oldest age group that Toppino and DeMesquita studied (10-year-olds) and older groups, as studied by Wilson. Finally, to allow a direct test of Wilson's theory that the crucial factor influencing the spacing effect is WM capacity, each child's digit span was measured.

Based on Wilson's theory, we would expect to replicate his results. The youngest children will remember more words at lag 1 than lag 0, but show little or no further benefit from longer lags. For adults, as the lag increases, so will their memory performance. The intermediate age groups will show a developmental progression between these two patterns. Furthermore, since the development of the lag effect is purported to result from the development of WM, the pattern of results should be similar or clearer when memory span is substituted for year group. That is to say, if the predicted pattern of association between lag and age does not emerge, this might be because of differences in memory capacity within the age groups. Therefore, it is predicted that those with the lowest WM capacity (a recorded digit span of up to four) will show a big increase in performance between lags 0 and 1, but little further benefit from longer lags, whereas those with higher WM capacity will show increased memory performance with increasing lag.

Method

Participants

A total of 119 participants took part in the experiment. Five age groups were tested, comprising approximately 24 participants from each year group. Details are given in Table 1. For each of the school year groups, students were arbitrarily selected from a mixed-ability class. The undergraduates were first-year psychology students at Warwick University, who received course credit in return for participation. Note that this group was

Table 1. Number of male and female participants, total number, and mean age, in each of the year groups

Group	n	Mean age (Yrs:Mths)	Male:Female
Year 1	23	5:6	12:11
Year 3	24	7:6	12:12
Year 6	24	10:7	12:12
Year 9	24	13:7	13:11
Undergraduate	24	19:4	19:5

beyond the age of compulsory education, potentially biasing the ability profile relative to the other groups.

Design

Two factors were initially included in the design of this experiment: Year group was a between-participants factor with five levels (Years 1, 3, 6, 9 and undergraduate) and lag was a within-participants factor, also with five levels (once presented items and repeated items at lags of 0, 1, 3 and 8 intervening items). Additionally, digit span was measured and included as a between-participants factor with five levels, chosen to equate, as far as possible, the numbers in each group (recorded spans of three and four, five, six, seven, and eight and nine).

Materials and apparatus

For the digit span test, three lists of digits were generated at each of eight list lengths from two through to nine digits. Each list was created by taking a random sample, without replacement, from the digits one to nine.

All words used in the experiment were drawn from a pool of 49 three- and four-letter, high frequency, concrete nouns. Five of these were used in a practice list, and 18 served as critical items in the main list with an additional three words as a primacy buffer and five words as a recency buffer. The final 18 words were used as foils in a final recognition test.

Firstly, a practice list was generated, consisting of five different words, of which two were presented once and one was presented at each of three different lags: 0, 3 and 6. The experimental list was 38 words long. The first three and last five positions were reserved for the primacy and recency buffers. The rest of the list was divided into three blocks of ten words each. Each block included one repeated word at each of the four lags (0, 1, 3 and 8 intervening items) and two once-presented words, which appeared in adjacent positions. Within each block, the first presentation of a lag 0 item appeared in position 2, the first lag 1 item was in position 4, the first lag 3 item in position 5 and the first lag 8 item in position 1. This list order was designed to minimize the possibility that serial position effects would interfere with the lag effect under investigation. An example of the experimental word list is given in the Appendix.

The 18 critical items were divided into six sets of three, and these were rotated through the different conditions. Since there were twice as many once-presented items as repeated items, each word appeared twice as often as a once-presented item than it did in any of the other conditions, but otherwise each word appeared equally often at each of the four different lags.

Finally, a list of words was created for a recognition test. This consisted of the 18 critical items and an additional 18 foils. These were presented in one of two random orders, one being the reverse of the other.

The words were presented on a portable computer.

Procedure

Participants were tested individually, away from the rest of their class. The first test to be administered was the digit span test. This was conducted according to the procedure described by Gathercole (1995), as follows. Starting with a length of two, two lists at each length were presented to participants. Following the presentation of a list, the participant repeated the list back without changing the order of the digits. Length was increased until a participant failed to repeat two lists of a given length; a third list of a given length was presented when just one of the first two at that length was recalled correctly. Span was defined as the longest length at which two lists were correctly recalled.

Following the digit span test, participants were instructed that they would see a list of words presented one at a time on the computer screen and that they should read out the words as they appeared and then, when the list had finished, report as many of the words as they could remember. They were told that many of the words would appear twice. The presentation rate was 5 s per word. After the practice list was presented, the participant was given as much time as they wished for oral free recall (answers were written down by the experimenter). At this point the instructions were clarified, if necessary, and the participant was warned that the following list was lengthy. The main word list was then presented, followed by a recall test. Both recall tests were terminated when the participant stated that s/he was unable to think of any more words, or looked uncomfortable.

Finally, a recognition test was administered. The experimenter read out a list of words, which included all the critical words from the main list mixed randomly with an equal number of foils, and the participant responded to each word with ‘Yes’ or ‘No’ as to whether or not it had appeared on the list. Two orders of the list were used; one was the reverse of the other. They were selected arbitrarily such that each was used for approximately half of the participants.

Results

For each participant, a digit span was recorded as well as the number of items in each condition (once-presented and lags 0–8) correctly recalled and recognized. Table 2 summarizes the total scores on these three measures for each year group. The recognition scores were close to ceiling for all conditions apart from the once-presented words, and were not analysed further.

The recall scores at different lags for each of the year groups are plotted in Figure 1. An alpha level of 0.05 was used for all statistical tests reported in this paper. A five (year group) × four (lag, excluding once presented items) mixed factors analysis of variance (ANOVA) test confirmed the differences apparent in Figure 1: There were significant effects of both year group ($F(4, 114) = 44.16$, $p < 0.001$) and lag ($F(3, 342) = 10.04$, $p < 0.001$). However, contrary to our prediction, there was no significant interaction between these factors ($F(12, 342) = 0.88$, $p = 0.57$), indicating that the deviations from the main pattern apparent in the graph probably do not represent systematic variations in performance. The one deviation that affects the pattern qualitatively, the lower recall score of Year 6 children for lag 8 than for lag 3 items, was further tested by repeating the analysis

Table 2. Total scores on the three tests for each of the year groups

Year group	Digit span		Total recall		Total recognition	
	M	*SD*	*M*	*SD*	*M*	*SD*
1	4.7	1.18	2.3	1.42	14.1	2.70
3	4.9	0.68	3.0	1.18	15.7	1.92
6	5.4	1.28	4.8	2.55	14.6	2.83
9	6.1	1.59	6.6	2.54	15.0	2.16
Undergraduate	7.6	0.88	9.8	3.19	15.5	3.91

Note: The maximum value for both recall and recognition scores was 18.

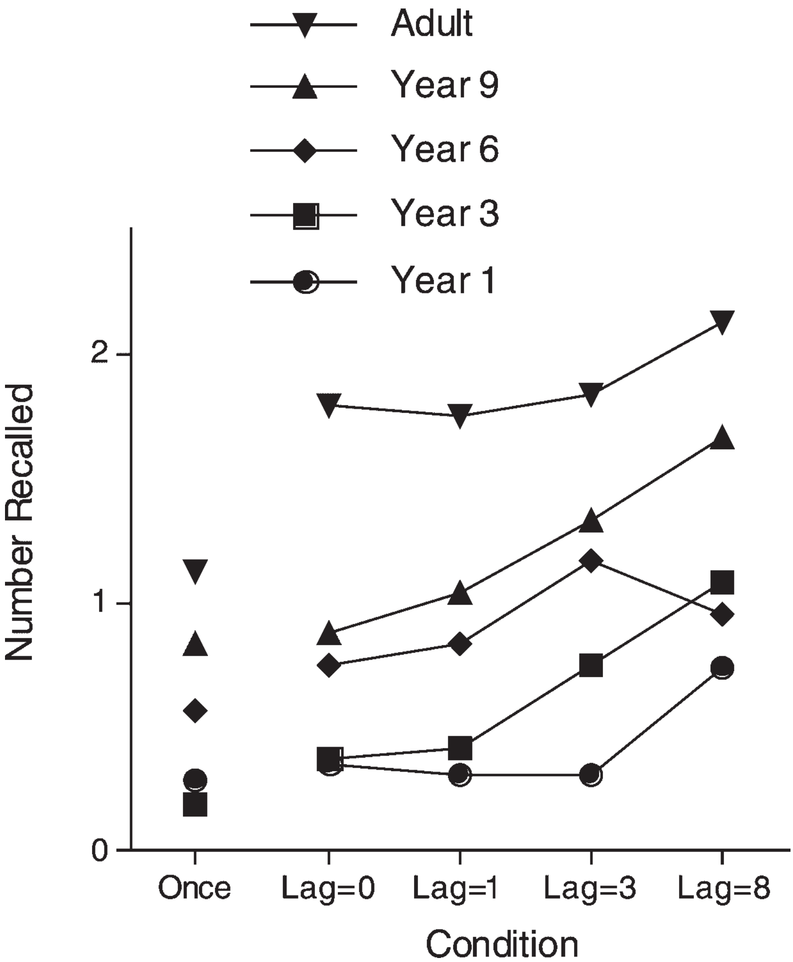

Figure 1. Recall scores (maximum 3) for each year group across different presentation lags

including only those lags. This found the same results as the main analysis; main effects of both year group ($F(4, 114) = 21.70$, $p < 0.001$) and lag ($F(1, 114) = 5.39$, $p = 0.02$) but no significant interaction ($F(4, 114) = 1.25$, $p = 0.295$). A second concern, from inspection of Figure 1, is the possibility of a floor effect with the youngest children. However, this should not concern us too much, as the most important difference is between lag 3 and lag 8. Wilson (1976) predicted that for young children there would be a benefit of short lags over massed presentation, but no further benefit of longer lags. In contrast, the youngest group, like the older children and the adults, show greater recall for lag 8 than lag 3. This indicates that the lag effect does not differ across different age groups.

A further ANOVA test was conducted with digit span (five levels; spans of three and four, n = 31; five, n = 28; six, n = 22; seven, n = 17; eight and nine, n = 21) substituted for year group (due to the high correlation between these factors, it is not appropriate to include both as independent variables in a single analysis). The results were similar to those obtained with year group: There were main effects of both digit span ($F(4, 114) = 22.2$, $p < 0.001$) and lag ($F(3, 342) = 8.86$, $p < 0.001$) but again, no significant interaction ($F(12, 342) = 0.97$, $p = 0.48$). The lag effect is the same, independent of WM capacity.

Discussion

The results of this experiment differ from those of Wilson (1976) in that all age groups benefited from increasing lags between presented items and there was no difference between age groups in the extent to which recall was affected by lag. In particular, the difference between the longest two lags (three and eight intervening items) was examined. A significant benefit of increased lag was found with, again, no interaction with age. This is in contrast to Toppino and DeMesquita's (1984) lack of a significant difference between these two lags for the 5- to 9-year-olds in their study. Given that there was a (non-significant) difference apparent in their results, it seems likely, in the light of the results of the current experiment, that Toppino and DeMesquita did not have a large enough number of participants (18 in each group, as compared with 24 in this experiment) to detect this difference and that their results are consistent with ours.

Furthermore, the direct test of Wilson's theory that WM capacity is critical to the spacing effect found no interaction between memory span and the effect of lag on recall. Wilson hypothesized that an increase in the spacing between presentations of an item improves subsequent recall because it increases the chance that the first presentation will no longer be in WM by the second presentation. Consequently, additional processing is needed to encode it a second time, resulting in a richer representation of the (twice encoded) item in long-term memory which in turn increases the likelihood of subsequent recall. Therefore, he predicted that participants with a short WM span (such as young children) would benefit most from short lags, but would not benefit so much from further increases in lag, relative to those with a longer WM span (such as adults), who would continue to benefit from increasing lags between items.

The current experiment has failed to support Wilson's hypothesis that the spacing effect is critically dependent on memory capacity. Instead, the results suggest that spacing item presentations is equally important to memory for adults and children of all ages, regardless of WM span. Therefore, these data rule out any theory of the spacing effect involving processes that develop with age, including both Wilson's (1976) WM hypothesis and Toppino and DeMesquita's (1984) encoding strategies theory. They do not rule out all versions of the variable encoding theory (but see Ross & Landauer, 1978, for evidence against this theory). From a practical point of view, the implication is that increasing the spacing of presentations will be equally beneficial to all ages. This is considerably more valuable than Wilson's suggestion that with young children, avoiding immediate repetition is important, but increases in spacing beyond that will yield no further benefit.

EXPERIMENT 2

Although Experiment 1 demonstrated the value of distributed over massed presentation of material for all ages tested, it may be argued that this comparison is not directly relevant to educational practice. Teachers already distribute material: Learning to read, for example, is distributed over many years and even much smaller topics will be covered over a week or more. Therefore massed presentation may not be the most appropriate control condition in experiments of the type reported here.

An alternative comparison is between different degrees of distribution, i.e. between few long sessions of teaching and many short sessions (the latter being more highly distributed). We refer to the former as 'clustered' presentation. This is not intended to identify a theoretically important presentation schedule but merely as shorthand for 'less distributed than the other schedule in the experiment.' In contrast to comparisons between massed and distributed practice, there is little difference in the overall duration of the schedules; the difference lies in the arrangement of practice within that time period. The remainder of this paper reports comparisons of clustered with distributed presentations, first in a laboratory setting (Experiment 2) and then in the classroom (Experiment 3). Experiment 2 is a variation on Experiment 1, also using episodic memory for lists of words.

Method

Participants
Twenty children and 16 adults took part in this experiment. The children were in Year 2 and had a mean age of 6 years 10 months. Most of the adults were first-year undergraduates taking part in return for course credit; a few were third-year or postgraduate students. None of the participants had taken part in the previous experiment.

Design
The experiment had two factors: Spacing, a within participants factor with three levels, massed, clustered and distributed, and age, a between participants factor with two levels, children and adults.

Materials

The words used in the experiment were drawn from the same pool as used in Experiment 1. A practice list consisted of four different words, which represented each of the different conditions: One was presented four times consecutively (massed), one twice consecutively (half of clustered), one twice at lag 4 (half of distributed) and one was presented once only (filler items). The main list was 40 words long. As with Experiment 1, the first three and last five items were primacy and recency buffers. The rest of the list was divided into two blocks of 16 words each. Each block included one word presented four times consecutively (massed), one word presented twice consecutively followed by eight intervening items then a further two consecutive presentations (clustered), one word presented four times with four intervening items between each presentation (distributed) and four words presented once each (filler items). It was necessary to present each word (apart from fillers) four times to create the clustered presentations condition. In order to keep the list sufficiently short for use with children, it was possible to include only two words in each condition. As with Experiment 1, the use of different words in each condition was counterbalanced.

Procedure
The procedure was the same as for Experiment 1, except that the digit span and recognition tests were excluded.

Results

The mean numbers of items recalled in the six conditions are shown in Table 3. A two-way ANOVA test found main effects of both schedule ($F(3, 102) = 14.55$, $p < 0.001$) and age

Table 3. Number of items recalled (maximum of 2) by children and adults in three learning conditions

Age group	Massed		Clustered		Distributed	
	M	*SD*	*M*	*SD*	*M*	*SD*
Children	0.70	0.54	0.70	0.43	1.15	0.45
Adults	1.06	0.53	1.25	0.63	1.88	0.94

($F(1, 34) = 28.10$, $p < 0.001$) but no interaction between them ($F(3, 102) = 0.57$, $p = 0.64$). Planned comparisons showed that distributed presentations resulted in better performance than either massed ($t(35) = 4.21$, $p < 0.001$) or clustered presentations ($t(35) = 3.75$, $p < 0.001$) but that there was no significant difference between the latter two conditions ($t(35) = 0.45$, $p = 0.65$).

Discussion

Clustered presentations produced recall performance that was no better than that following massed presentations. Distributed presentations resulted in significantly better recall than seen in either of the other two conditions. This was true for both children and adults, with no significant difference in the pattern of results between the two age groups. The difference between highly distributed teaching sessions and more clustered sessions is exemplified by two instructional programmes for teaching reading currently in use in the UK. The first of these is the National Literacy Strategy (NLS) and the second is the Early Reading Research project (ERR), which is being investigated in a number of schools in Essex (Solity, 2000). Whereas the NLS requires literacy to be taught in a single, hour-long session each day, teachers taking part in the ERR project teach reading in three 12-min sessions per day. Therefore, the NLS might be considered to use clustered presentations, with just one teaching session per day, relative to the ERR's use of presentations that are distributed throughout each day. With this in mind, Experiment 3 was designed to investigate the effect of scheduling teaching within the classroom.

EXPERIMENT 3

Experiment 2 established that, in a laboratory setting, distributed presentations of words resulted in better memory for those words than did clustered presentations. This leads us to expect that in a classroom setting, many short teaching sessions would produce better learning than fewer, longer sessions. This has particular relevance in the light of the NLS, which provides all primary school teachers (up to Year 6; 11-year-olds) with guidelines on teaching literacy. This includes strong emphasis on the requirement that literacy be taught in a single session each day, known as the literacy hour. In contrast, an alternative reading framework, the ERR project, has been developed that requires reading to be taught in three 12-min sessions per day. Children taught according to the ERR strategy have substantially out-performed those taught within the NLS (Solity, 2000).

The effect of varying the distribution of teaching sessions (isolated from other differences between ERR and NLS) was investigated in the final experiment, in which teachers were asked to make a small modification to their usual teaching of reading skills. Massed presentations of materials (i.e. all presentations of an item being given in a single session) were not included in this experiment for two reasons. Firstly, and most importantly, massed presentation of material is rarely used by teachers, and so the comparison is not relevant to the practical application. Secondly, Experiment 2 demonstrated that clustered presentations conferred no advantage over massed, so there is little value in including massed presentations for comparison when clustered presentations are the most appropriate control.

The elements of reading that were being taught in this experiment were correspondences between phonemes and single letter graphemes that represented them (i.e. letter sound correspondences), between phonemes and letter combinations that represented them (where one phoneme maps on to one grapheme which contains two or more letters, for example, 'ch'; 'ai'; 'er'; etc.) and reading of phonically regular words (where each phoneme in the word is represented by a single letter grapheme e.g. sit, bend, flip, stand etc.). Since the classes being taught were of mixed ability, none of these three tasks would have been appropriate to all of the children, but by including all three in our dependent measure, we cover tasks on which each child is progressing at the time of the study.

Method

Participants

Thirty-four Year 1 children (mean age 5 years 6 months) from two similar schools took part in this experiment. Twelve children attended one school and 22 attended the other (note that both classes included other children who were present during teaching sessions but were not available for testing; the classes were of a similar size). Both schools were participating in the ERR scheme. The implications of this that are relevant to this study are that both teachers were following the same procedures to teach reading and that reading was taught in three sessions per day.

Whilst it might be argued that it would have been preferable to compare different teaching methods executed by the same teacher, it should be borne in mind that this does not really remove teacher differences since any teacher will have different attitudes towards the different teaching methods. Furthermore, there are other problems associated with such a design: If similar materials are used for the two conditions there will be carry-over effects from one teaching method to the next; if different materials are used it is almost impossible to make a meaningful comparison. Therefore, it was decided to compare two teachers who were using the same, precisely defined, method of teaching.

Design

A between participants design was used to compare teaching for three 2-min sessions per day ('distributed') with teaching for one 6-min session per day ('clustered'). The dependent variable was the improvement in a combined score of letter-sound knowledge, letter-combination knowledge and reading of phonically regular words ('phonics').

Materials and teaching method

Children were taught to associate each of the letters with its sound (score out of 26), to read words including the following letter combinations: ck, th, er, ing, sh, ch, wh, qu, ar and ea (score out of 40) and to read three-, four- and five-letter phonically regular words (score out of 20). For the purpose of establishing overall reading ability, children were tested on sight vocabulary (reading up to 100 common words) and skill in synthesis (blending sounds) and segmentation (separating sounds; both scored out of 20). The procedure for teaching words containing newly-learnt grapheme-phoneme correspondences was as follows. First the teacher demonstrated how to read the word, sounding out each phoneme before saying the whole word. Then the children and teacher repeated this together and finally the children repeated this without the teacher. Each teaching session included some familiar items, which were tested only (the children read the word without help from the teacher), some newer items, and a set of words in which new and old items were interspersed. The phrases 'new items' and 'familiar items' refer to grapheme-phoneme correspondences and the structure of word, not to actual words presented. For example, in learning the 'sh' sound, children would read a number of different words that all included 'sh'. Over the period of this experiment, although the actual words presented from session to session varied, very few new items were introduced, so essentially the same material was repeated in each session of the experiment.

Procedure

Children were initially tested, individually, on all six reading skills. This was followed by 2 weeks of teaching phonics, one class using distributed presentations and the other using clustered presentations. This was integrated into the usual teaching of reading, with the grouping of the three 2-min sessions per day into a single session for clustered presentations being the only change to usual teaching practice. Since both teachers were following the ERR scheme, they were both following the procedure described above. They were selected for this experiment on the basis of their close adherence to the scheme after monitoring. Therefore any non-experimental difference between their teaching procedures was kept to an absolute minimum.

After 2 weeks of teaching, children were tested again on letter sounds, letter combinations and reading of phonically regular words ('phonics').

Results

Firstly, correlations were calculated between all the time 1 measurements. These are shown in Table 4. It is clear that all four measures are highly intercorrelated, so a principal components analysis was conducted on these measures (as a means of data reduction, not to test for underlying factors). This resulted in a single factor that accounted for 84% of the variance and was heavily weighted on all four variables. This factor was used as a measure of overall reading performance in the following analysis.

The two groups of children were compared on the combined reading measure (time 1) and found to be significantly different ($t(32) = 3.06$, $p < 0.01$). Inspection of the data revealed that the larger group included a wider range, overlapping with those of the

Table 4. Pearson's correlations between time 1 reading measures

Measure	1	2	3	4
1. Phonics	—	0.892*	0.748*	0.803*
2. Sight vocab.		—	0.788*	0.733*
3. Synthesis			—	0.814*
4. Segmentation				—

*$p < 0.001$.

smaller group, which included only higher performing children. Therefore, children in the larger group whose reading scores fell outside the range of the smaller group were excluded. This left 12 children in each group. A further *t* test confirmed that the revised groups were matched on reading performance scores ($t(22) = 0.92$, $p = 0.37$). Therefore, the resultant two groups were of the same size and had comparable baseline ability.

The groups were then compared on improvement scores in phonics (score at final test minus score at initial test on the 'phonics' tasks, not including sight vocabulary, synthesis or segmentation) to assess the two teaching schedules. Children who had received distributed teaching sessions showed significantly more improvement than those who had received clustered teaching sessions (8.3 vs. 1.3 points improvement; $t(22) = 3.05$, $p < 0.01$).

Discussion

This experiment demonstrates that applying the results of laboratory studies (e.g. Experiment 2) to classroom practice can have a significant impact on learning. Over 2 weeks, children whose teaching consisted of three 2-min sessions per day showed more than six times the improvement of those who were taught for one 6-min session per day.

This experiment included no delayed test (the delay being only 1 day, relative to a teaching period of 2 weeks). However, a delayed test would be unlikely to weaken the observed difference, since the benefit of distributed practice tends to increase with delay (Glenberg, 1979). Secondly, in a field such as reading, it is unrealistic to talk of memory surviving a delay, since the task is ubiquitous.

GENERAL DISCUSSION

There are three main findings reported in this paper. The first is that children of all ages show the same effect of increasing lag between items as adults do and this effect is not related to WM span. The second is that even with the more ecologically valid control condition of clustered presentations, there is still a strong benefit of distributed practice. Thirdly, these effects can be applied in the classroom, such that small modifications to teaching practice can produce substantial improvements in learning.

Experiment 1 demonstrated that the lag effect is similar in magnitude over a wide range of ages. This finding has theoretical as well as practical implications. If the lag effect does not develop with age then it cannot depend on a process that does, such as WM (also tested directly in Experiment 1) or organizational strategies, as suggested by Toppino and DeMesquita (1984) following Glenberg (1977).

For the spacing effect to be valuable in education, it is necessary to test distributed practice against a more relevant control condition than massed practice. To this end, Experiment 2 introduced clustered presentations, representing few, long sessions of teaching as compared with many, shorter sessions. This demonstrated the value of distributed over clustered presentations in a laboratory setting, with both adults and children. In fact, for children, massed and clustered presentations resulted in identical performance levels.

Extending the laboratory-based findings to classroom practice, Experiment 3 demonstrated the value of distributed presentations in teaching when compared with clustered presentations. This experiment has strong ecological validity, since the usual teaching of reading was used as the basis of the experiment. Two teachers using the same teaching methods (according to the ERR project) were compared, after asking one of them to modify only the scheduling of teaching sessions. In this experiment a large effect was obtained, with improvement scores following highly distributed teaching being more than six times the improvement scores that followed the less distributed sessions. Whilst this study was small, and hence of limited scope, it illustrates the potential of applying distributed practice in the classroom, and lays the groundwork for more extensive investigations into the application of this phenomenon in education.

The clearly prescribed teaching procedures associated with the ERR created the opportunity to run an experiment within a school context where experienced teachers, rather than researchers, implemented the intervention. This potentially gives the experiment considerable credibility with teachers, demonstrating that viable and realistic classroom practice can emerge from well controlled, laboratory experiments.

The outcomes of this research have considerable implications for how learning is organized for children. Typically they are taught core literacy skills in single teaching sessions and this is now standard practice in the majority of English schools since the introduction of the NLS, where children are taught for a concentrated block of an hour a day. Breaking up this hour into shorter periods distributed throughout the day, particularly when children are learning sight vocabulary, phonological and phonic skills, could have an enormous impact on children's attainments without having any financial or resource implications for schools.

ACKNOWLEDGEMENTS

We thank Rachael Deavers for conducting Experiment 3.

REFERENCES

Cahill, A., & Toppino, T. C. (1993). Young children's recognition as a function of the spacing of repetitions and the type of study and test stimuli. *Bulletin of the Psychonomic Society, 31*, 481–484.

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 389–396.

Cornell, E. H. (1980). Distributed study facilitates infants' delayed recognition memory. *Memory & Cognition, 8*, 539–542.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162–170.

Dempster, F. N. (1988). The spacing effect: a case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627–634.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (2nd ed., pp. 317–344). London: Academic Press.

Gathercole, S. E. (1995). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology, 65*, 155–164.

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect of free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 3*, 282–294.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95–112.

Greene, R. L. (1989). Spacing effects in memory: evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 371–377.

Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1004–1011.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596–606.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning, 4*, 11–18.

Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition, 15*, 436–443.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition, 10*, 97–102.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308.

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17*, 669–680.

Solity, J. E. (2000). The Early Reading Research: applying psychology to classroom practice. *Educational and Child Psychology, 17*, 46–65.

Toppino, T. C. (1991). The spacing effect in young children's free recall: support for automatic process explanations. *Memory & Cognition, 19*, 159–167.

Toppino, T. C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society, 31*, 27–30.

Toppino, T. C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology, 37*, 637–648.

Toppino, T. C., Kasserman, J. E., & Mracek, W. A. (1991). The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of Experimental Child Psychology, 51*, 123–138.

Wilson, W. P. (1976). Developmental changes in the lag effect: an encoding hypothesis for repeated word recall. *Journal of Experimental Child Psychology, 22*, 113–122.

APPENDIX

Example of a word-list used in Experiment 1.

	hat	Primacy buffer
	gold	
	door	
lag 8	rock	Block 1
lag 0	snow	
lag 0	snow	
lag 1	book	
lag 3	soil	
lag 1	book	
once presented	cake	
once presented	sun	
lag 3	soil	
lag 8	rock	
	man	Block 2
	boat	
	boat	
	ball	
	box	
	ball	
	hill	
	car	
	box	
	man	
	farm	Block 3
	town	
	town	
	bed	
	wall	
	bed	
	park	
	key	
	wall	
	farm	
	tree	Recency buffer
	step	
	rose	
	star	
	room	