

The longer-term effects of reading interventions delivered by experienced teaching assistants

ROBERT SAVAGE, SUE CARLESS and OZLEM ERTEN

This study explored the responsiveness of children at risk of reading problems in Year 1 to a phonics intervention delivered by teaching assistants (TAs). Based on their non-word decoding skills in the immediate post-tests, 74 children were clustered together at the high end as 'treatment responders' ($n = 49$) and at the low end as 'treatment non-responders' ($n = 25$) and were followed up at the end of Key Stage 1, 16 months after the intervention finished. The treatment-responder group was superior in all areas of rated attainment and, unlike the non-responders, achieved national averages in most teacher ratings of attainment. These results suggest that experienced TAs can help two out of three children at risk of reading difficulties.

Key words: reading intervention, teaching assistants.

Much research shows that early phonic interventions are effective when they focus on operations such as segmenting or blending and phoneme-grapheme instruction (for example, Ehri *et al.*, 2001; Savage and Carless, 2008). A key issue is whether reading interventions have long-term impacts on the performance of children. An analysis of five intervention studies conducted by Torgesen (2000) showed that around 60%–80% of children initially reading below the 20th percentile were able to achieve word-reading skills almost above the 30th percentile at post-intervention follow-up after phonics-based interventions.

Given that controlled research on interventions in phonics training has generally proved to be *efficacious* for early literacy development, it is crucial to continue examining the most efficient ways of developing *effective* interventions that will operate well in real-world settings of schools (for example, Savage and Carless, 2004, 2005, 2008). There is relatively little of this latter sort of evidence. Evidence from a series of phonemic awareness training interventions in the UK has strongly suggested that teaching assistants (TAs) can have significant impacts on children's early word

reading, letter-sound knowledge, phonological skills and pure decoding abilities (Savage *et al.*, 2003; Savage and Carless, 2005, 2008; Hatcher *et al.*, 2006).

Only one study to date however has explored longer-term effects of these interventions. Savage and Carless (2008) divided children into 'treatment responders' and 'treatment non-responders' based on their decoding skills. They found that the two groups differed at delayed follow-up, with the treatment-responder group more likely to reach national averages in reading. This study was limited in that the sample was relatively modest in size ($n = 55$ children in total). Secondly, the TAs in this study were trained for their role for the first time. It may be that with both further training and reflection, TAs may be even more effective in delivering the same interventions in the second year. These important issues concerning the practical effectiveness of TAs are thus explored here.

This study focused on two questions:

1. Is there raised attainment for treatment responders 16 months after the intervention at the end of Key Stage 1 compared to treatment non-responders?
2. Do treatment responders go on to reach average levels of attainment?

Method

The original intervention study was carried out in three phases: a pre-test screening phase, an intervention phase and an immediate post-test phase. Details follow those reported in Savage *et al.* (2003) unless stated here. Children were seven years of age at the point of the 16-month follow-up assessment. An overview of the design is presented in Figure 1.

Participants

Screening phase

In the screening phase TAs were taught to screen children on measures of phonological awareness, reading, spelling and

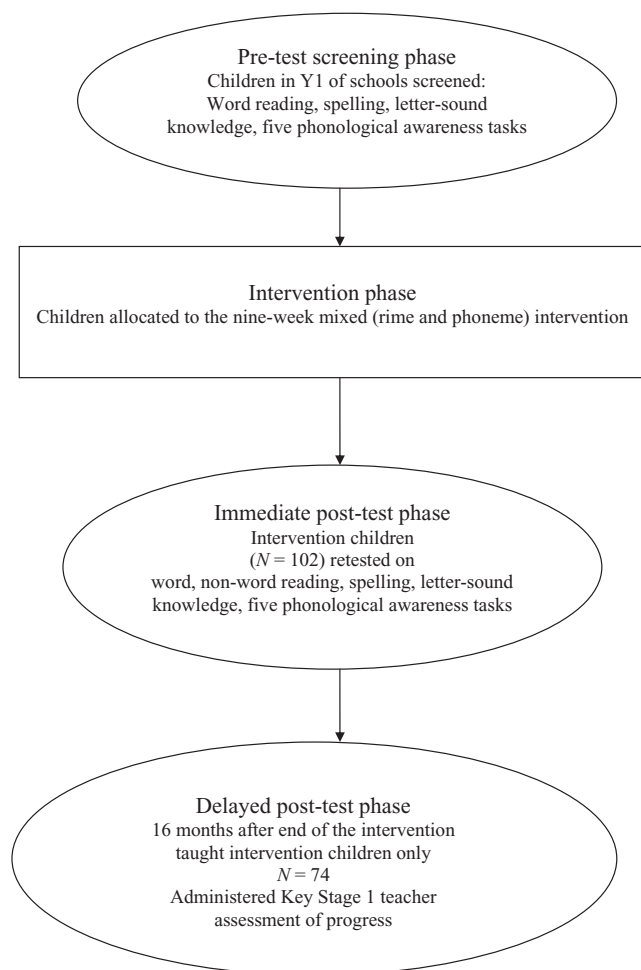


Figure 1. Overview of the intervention project

letter-sound knowledge in 12 classes. TAs screened those children identified by the regular class teachers as being of concern. Alongside advice from the second author, the TAs then chose the children showing the weakest performance across tasks for intervention. The mean pseudo-word decoding score at pre-test in the present sample was 0.24 ($SD = .72$), out of a possible maximum of 12. This was not significantly different from the pre-test mean of the 20% of children selected as the poorest readers from school-wide sampling in Savage *et al.* (2003), $t(74) < 1$, *n.s.*

Intervention phase and delayed post-test

Seventy-four children receiving TA-delivered intervention participated in all phases of the study. There were 39 boys and 35 girls. These were a subset of 102 children who initially received the intervention and for whom records were available 16 months after the intervention closed.

Materials

Literacy assessments at pre-test

Children were asked to read 12 pseudo-words (for example dat, tav, mip). Children were asked to read six words (at,

had, let, dig, cut, top) and asked to write six words (sat, hop, but, red, win, leg). Children were shown cards with 26 individual letters on them and asked to say the sound that goes with the letter. The Spearman-Brown internal reliability of each of these tests was .90 or above.

Baseline assessment

The Foundation Stage Profile (FSP, Hartas, 2006) is made up of 13 scales based on the Early Learning Goals (ELGs). This rating scale is completed by teachers based upon their observations of children in regular school settings.

Measures delivered at immediate post-test

All of the above measures except the FSP administered at pre-test were re-administered in the same fashion at the immediate post-test.

Measures administered at delayed post-test at age seven

At the end of Key Stage 1, the teacher's assessment of each child's current global attainment in English, Maths and Science was measured against the expected average performance of children in the National Curriculum. English is subdivided into Speaking and Listening, Reading and Writing sub-components. Teacher assessments are recorded as 'working towards Level 1 (W)', and Levels 1, 2 and 3 of the curriculum. Level 2 and above correspond to the average and above-average levels.

Procedure

In this study all children received the same 'mixed' rime-based 'analytic' and phoneme-based 'synthetic' programme described fully by Savage *et al.* (2003). The children were seen for 20 minutes three times per week. As an addition to our previous approaches, we gave TAs laminated 'caterpillar' phoneme frames with letter tiles of the ten letters Velcroed at the top. Each child had one of these frames on which they manipulated the CVC words. The TAs also made A3 laminated sheets with all the alphabet tiles Velcroed around the sheet so that, when confidence grew, the children could manipulate CVC words using more letters. They were also able to write on the laminated sheets.

Children were seen during the word-level work of the National Curriculum for the main part. In the first part of each session, children would all engage in letter-sound learning activities. Children used a range of multi-sensory approaches (for example, saying, looking, tracing) to learn letter sounds supported by the Jolly Phonics stories and actions (Lloyd, 1992). Only one vowel (a) and nine consonants were taught (initially) to reduce memory load and to move on quickly to teach the principle of segmenting and

blending with a limited number of letters. This work would be recapped at each new session.

TA training and re-training

The majority of TAs had participated in a previous intervention (Savage and Carless, 2008) and had at that point received a morning's training on running the intervention delivered by the second author (an experienced literacy resource specialist). In this second year of intervention, the TAs requested and received a further repeat training. TAs gathered together. The first morning's training was given over to revisiting the assessment materials administered by TAs at pre- and post-test and providing an overview of the programme. This enabled TAs to complete the assessment and then at the next session (training on the intervention) they came with any queries. The second author was able to give advice on grouping the children as a result of their assessments. On the second morning, the materials were revisited in precise detail. This provided an opportunity both to clarify and to discuss and link the use of the materials to experiences and scenarios encountered in the previous year's intervention. Those who were more experienced were a useful source of information in the training for the minority of TAs new to the project. Attention was also directed to the management of effective small groups of Year 1 children in this session. Members of the Local Authority Learning Support Team were always available in schools for consultation during the intervention, and some TAs did ask their advice.

Results

At the time of delayed post-testing, complete data for 74 children were available from a total of 102 seven-year-olds. The one exception to this is that in three cases Foundation Stage Profile data were not available. Analyses using this particular variable are thus based on $n = 71$ participants. The 74 children represented 73% of the original poor-reader sample. The follow-up sample consisted only of those children receiving the intervention and who either responded to – or did not respond to – this intervention. Preliminary analyses confirmed that the subset of 74 poor readers did not differ from the original set of 102 on important pupil background measures routinely collected by the local education authority: special educational need, free school meal eligibility or ethnicity or on word and pseudo-word reading, spelling, letter-sound knowledge or phonological skill ($F < 1$ *ns* in all cases).

Preliminary comparisons of treatment responders and non-responders

The categorisation of treatment responders and non-responders was based solely on pseudo-word decoding ability. We first explored the distribution of decoding skills depicted in Figure 2 to determine a cut off-point. Figure 2 shows that the scores exhibited characteristics of a bimodal

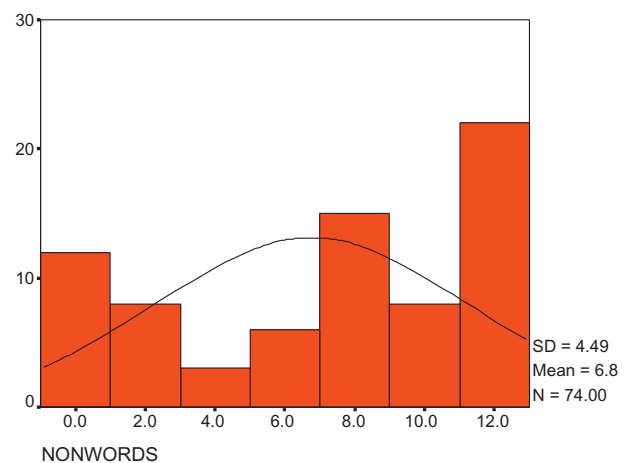


Figure 2. Distribution of the number of correctly decoded pseudo-words at the time of immediate post-testing

distribution: the two most frequent responses were eight and 12 as maximum scores and scores of four and six being the least frequent responses. We therefore undertook a median split of this distribution. Those children clustered at the high ends were identified as 'treatment responders' ($n = 49$) and those who were clustered at the low ends were identified as 'treatment non-responders' ($n = 25$).

Responders and non-responders were first formally contrasted on: Gender; Ethnicity; Special Educational Need (SEN, coded as those requiring additional 'school action plus' support); Free School Meals (FSM), as well as the FSP. These data are presented in Table 1. Of all children 94.6% had English as their mother tongue, and four children (5.4%) came from other ethnic backgrounds that were not considered as White UK or White European which is broadly typical of the UK. Boys did not drastically outnumber girls in this sample, making up 52.7% of the sample (girls = 47.3%). The proportion of children eligible for Free School Meals, a measure of socio-economic status, was 23%. This proportion did not differ from either national or local norms for FSM eligibility, $\chi^2 (1, N = 74) < 1$, *ns*.

The pupil background data on Gender, Ethnicity, SEN and eligibility for FSM were analysed using χ^2 and the data on the Foundation Scale Profile were analysed through *t*-tests. These analyses revealed no significant differences between treatment responders and non-responders, $\chi^2 (1, N = 74) < 1$, *ns* in all cases except for SEN where more children in the treatment non-responder group were considered to require additional support (school action plus or Statement), $\chi^2 (1, N = 74) = 7.94$, $p < .01$, and the performance of the children on the Foundation Stage Profile Baseline Assessment which showed that the treatment non-responder group scored lower on this measure than the treatment responder group, $t (70) = -2.83$, $p < .01$.

Follow-up of treatment responder versus non-responder groups at end of Key Stage 1 testing

Data on Key Stage 1 attainment in teacher assessment are presented in Table 2. Visual inspection of these data shows

Measure	Gender	Ethnic	SEN	FSM	MTO	FSP
Treatment responders	F = 24 M = 25	W = 41 O = 7 R = 1	0 = 42 1 = 7	1 = 9 2 = 40	E = 47 O = 2	71.09 (15.57)
Treatment non-responders	F = 11 M = 14	W = 20 O = 5 R = 0	0 = 14 1 = 11	1 = 8 2 = 17	E = 23 O = 2	60.08 (16.04)

Key:

Gender Gender of sample by number of males (M) and number of females (F)

Ethnic Ethnicity of sample (W = White UK or White European, O = any other background, R = refused to classify self)

SEN Number of children identified as experiencing special educational needs of sufficient severity to require external specialist support (1)

FSM Number of children eligible for free school meals (1 = eligible; 2 = not eligible)

MTO Mother tongue of children (E = English; O = Other than English)

FSP Foundation Stage Profile

Table 1. Pupil background and baseline assessment task data for treatment responders and non-responders

quite marked differences between the number of children achieving a Level 2 (the national average for Key Stage 1 attainment) in the teacher assessment of attainment among treatment responders and non-responders. Non-responders were more likely to achieve the below-average Level 1 designation. All these results were analysed formally using χ^2 . For the purpose of these analyses we collated all scores below Level 2 (below national average attainment) and contrasted scores at Level 2 or above (achieved national average attainment) in order to obtain reasonable sample sizes in all cells. These analyses revealed significant differences between treatment responders and non-responders on teacher assessment of Speaking and Listening, χ^2 (1, $N = 74$) = 11.50, $p < .001$; Reading, χ^2 (1, $N = 74$) = 17.17, $p < .001$; Writing, χ^2 (1, $N = 74$) = 10.32, $p < .001$; and Science, χ^2 (1, $N = 74$) = 18.22, $p < .001$; but not for Mathematics, χ^2 (1, $N = 74$) = < 1, *ns*.

Table 2. Percentages of the performance of children who are treatment responders and non-responders on Key Stage 1 teacher assessments

National Attainment Level	X	W	1	2	3
1. Speaking and Listening					
Treatment responders	0	0	12	88	0
Treatment non-responders	0	8	40	52	0
National norms	0	2	10	65	23
2. Reading					
Treatment responders	0	0	16	80	4
Treatment non-responders	0	8	56	36	0
National norms	0	3	11	59	27
3. Writing					
Treatment responders	0	4	29	65	2
Treatment non-responders	0	16	56	28	0
National norms	0	5	13	67	15
4. Mathematics					
Treatment responders	0	0	14	84	2
Treatment non-responders	0	0	28	72	0
National norms	0	2	7	68	23
5. Science					
Treatment responders	0	0	4	96	0
Treatment non-responders	0	0	44	56	0
National norms	0	2	8	65	25

Key:

X Children excluded from test or otherwise unavailable

W Working towards level

1 Level 1 in the National Curriculum achieved

2 Level 2 in the National Curriculum achieved

3 Level 3 in the National Curriculum achieved

Analyses controlling for the influence of extraneous variables FSP, SEN

Treatment-responder and treatment non-responder groups differed on their FSP scores and in the number of children with identified special educational needs. In order first to control for the effects of the FSP score and of SEN coding to allow an exploration of the independent link between treatment responsiveness and attainment, a series of sequential binomial logistic regressions were run. Binomial logistic regressions analyses can explore the combined and independent effects of multiple independent variables on a given DV. In this way the influence of extraneous factors (such as SEN and FSP in this case) on attainment can first be evaluated at steps 1 and 2 of such analyses before the *independent* effects of treatment responsiveness and attainment are considered at step 3. Binomial logistic regression is a powerful technique here as it can explore binary DVs (responder versus non-responder), and as a non-parametric technique does not need to make assumptions about the distribution of variables (Tabachnick and Fidell, 2001). Preliminary analyses confirmed that only around 10% of variance was shared by FSP, SEN and Response to Intervention (RTI) variables. Standard techniques also revealed no outliers in variables.

Preliminary analyses also confirmed that in all cases there was appropriate goodness of fit in these analyses: appropriately designed stepwise logistic regressions should exhibit non-significant χ^2 tests for goodness of fit. As shown in Table 3 all of these analyses were non-significant.

In logistic regression analyses, SEN status and FSP score were entered at steps 1 and 2, respectively, and the dichotomous variable of Treatment Responsiveness (treatment responder versus treatment non-responder) was then entered at step 3 of regression analysis. In each of five stepwise regression analyses reported, a different measure of attainment (Speaking and Listening, Reading, Writing, Mathematics, Science, respectively), operated as the dependent variable. The results of these analyses and the main logistic regression analyses are presented in Table 3.

Inspection of Table 3 confirms that across attainment measures both FSP and SEN have impacts on attainment at Key

Table 3. Predictors of Key Stage 1 Literacy Assessments

DVs	Speaking and Listening		Reading		Writing		Mathematics		Science	
IVs	Wald <i>F</i>	<i>df</i>	Wald <i>F</i>	<i>df</i>	Wald <i>F</i>	<i>df</i>	Wald <i>F</i>	<i>df</i>	Wald <i>F</i>	<i>df</i>
SEN	2.94	1	3.40	1	.92	1	8.95**	1	4.64*	1
FSP	2.93	1	1.00	1	2.66	1	3.13	1	3.16	1
Treatment	4.33*	1	8.70*	1	5.16*	1	.01	1	7.51**	1
H-L fit	<i>p</i> = .18		<i>p</i> = .39		<i>p</i> = .99		<i>p</i> = .46		<i>p</i> = .97	

Note: * = *p* < .05, ** = *p* < .01.

Key:

Df Degrees of Freedom

SEN Special Educational Need

FSP Foundation Stage Profile

Treatment Treatment responsiveness

H-L fit Hosmer-Lemeshow Goodness of Fit measure

Stage 1 when entered at steps 1 and 2. These effects were statistically significant for Mathematics (FSP and SEN) and in Science (for SEN only), whereas non-significant trends were evident for these measures for Speaking and Listening, Reading and Writing. The regression analyses in Table 3 also show that after controlling for the effects of FSP and SEN there were still significant effects of Treatment Responsiveness on teacher assessment of Speaking and Listening, Wald $F(1, N = 72) = 4.33, p < .05$; Reading, Wald $F(1, N = 72) = 8.70, p < .01$; Writing, Wald $F(1, N = 72) = 5.16, p < .05$; and Science, Wald $F(1, N = 72) = 7.51, p < .01$; but not for Mathematics, Wald $F(1, N = 72) = < 1, ns$. The unique proportion of variance explained by Treatment Responsiveness at step 3 was: 7% for Speaking and Listening, 13% for Reading, 8% for Writing, 0% for Mathematics and 16% for Science.

Follow-up of treatment responder versus national average performance at Key Stage 1

Data on the national norms are presented in the lower section of Table 2. Comparison of treatment responders versus national norms for teacher assessment showed that treatment responders did not differ from national averages in Speaking and Listening, $\chi^2(1, N = 74) = 0, ns$; Reading, $\chi^2(1, N = 74) < 1, ns$; Mathematics, $\chi^2(1, N = 74) = 1.23, ns$ and Science, $\chi^2(1, N = 74) = 2.75, ns$. The treatment responders did not however reach national averages in Writing, $\chi^2(1, N = 74) = 5.92, p < .05$. Comparison of treatment non-responders versus national norms for teacher assessment showed that treatment non-responders always scored lower than the national averages: Speaking and Listening, $\chi^2(1, N = 74) = 30.86, p < .001$; Reading, $\chi^2(1, N = 74) < 45.44, p < .001$; Writing, $\chi^2(1, N = 74) = 58.90, p < .001$; Mathematics, $\chi^2(1, N = 74) = 11.97, p < .001$; and Science, $\chi^2(1, N = 74) = 29.32, p < .001$.

Discussion

This article explored two questions:

1. Is there raised attainment for treatment responders 16

months after the intervention at the end of Key Stage 1 compared to treatment non-responders?

2. Do treatment responders go on to reach average levels of attainment?

Our two key findings are firstly that children who responded to interventions were clearly distinguishable from those who did not respond strongly to interventions, based on teacher assessment of attainment at Key Stage 1, some 16 months after the reading intervention had finished. Secondly, around two-thirds of children initially at risk who showed signs of responding to interventions later achieved national average performance on teacher attainments at the Key Stage 1 level in all domains except for Writing. Such group-specific improvement effects cannot be explained by general statistical artefacts such as regression to the mean as this would tend to produce either equivalent performance across treatment response groups or advantages for the non-responder group which started with poorer performance on several measures including pseudo-word decoding.

Importantly, our analyses show that while the children here did differ in terms of their baseline assessment and levels of special educational needs at school, baseline skills and special educational needs cannot provide an explanation of results reported here as the regression analyses reported in Table 3 explicitly controlled for baseline skills and special educational needs and still showed significant effect of treatment group (responder versus non-responder) on Key Stage 1 test results *after* the influence of these other variables on Key Stage 1 performance was considered. It is also worth noting that the effect sizes for the *unique impact* of the TA-delivered intervention were reasonably large, explaining between 7% and 13% of the variance in literacy outcome measures.

On the other hand, about a third of this at-risk sample did not respond to the TA-delivered intervention by showing abstract decoding skills, and as a whole this group was clearly at significant risk of further reading problems and general academic difficulties. Many of these children may require ongoing literacy support and differentiation to achieve national average levels in their school careers. This tail of children with continuing at-risk profiles represented

33% of the original lowest-performing 20% at initial full-school screening. Thus, if these figures are generalisable, then about 7% of children will require further and possibly ongoing support in addition to both effective early whole-class teaching and TA-led early phonics intervention.

These data thus appear to provide theoretical and practical evidence on the importance of TA-delivered interventions in order to meet the needs of many children at risk of reading difficulties and they replicate and extend the work of Savage and Carless (2008) with a new and somewhat larger sample. Our findings here predict that a majority of a second wave of children would reach national averages at follow-up in teacher attainment scores of Key Stage 1. This assumption should however be verified with further evidence from larger nationally representative samples.

The finding that two out of three children reached at least 50% success in nonsense-word decoding represents a sharp improvement on the previous year. In Savage and Carless (2008) only 15 out of 55 (27%) of at-risk children mastered the same non-word decoding task with even 40% accuracy. While some caution needs to be exercised in generalising across cohorts of children, this suggests that trained staff administering interventions as well as at-risk students might benefit from phases of training, retraining and support in their delivery of literacy interventions.

Limitations of the present study

One potential limitation of the present study is that it did not incorporate a formal contrast of the impact of TA-delivered interventions with a control condition where there was no such intervention for children. Our previous experimental studies run in the same schools with comparable cohorts of the poorest readers at the beginning of Year 1 have however shown that such subsets of children do not show *any* gain in the cognitively demanding abstract task of pseudo-word pronunciation over the nine weeks of the intervention (for example, Savage *et al.*, 2003). Children in the present study showed almost no pseudo-word decoding ability at pre-test, and were also in this way comparable to the samples reported in Savage *et al.* (2003) and Savage and Carless (2005, 2008), adding to our confidence that pseudo-word decoding at post-test measures response to intervention specifically.

Wider implications

This research provides further evidence to a body of studies implemented in the UK suggesting that early school-based reading interventions, delivered by TAs and supported by teams of professionals, can contribute significantly to improving reading outcomes for children over the short and medium term (Savage and Carless, 2008). The present findings may have important implications for maximising *cost-*

effectiveness of school-based interventions with respect to fostering beginning reading skills as the intervention was cheap to construct and deployed existing staff in preventative roles.

Acknowledgements

We would particularly like to thank all of the staff and parents of children in the London Borough of Sutton, and particularly the schools' teaching assistants who prepared and carried out the intervention work so enthusiastically and effectively. We would like to thank the J. J. Trust and the Helen Arkell Dyslexia Association for financial and professional support that made this project possible. We would particularly like to thank Anne Mitchell, formerly Assistant Director of the Helen Arkell Dyslexia Centre.

References

- EHRI, L. C., NUNES, S. R., WILLOWS, D. M., SCHUSTER, B. V., YAGHOUB-ZADEH, Z. and SHANAHAN, T. (2001) Phonemic awareness instruction helps children learn to read: evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250–287.
- HARTAS, D. (2006) *Dyslexia in the Early Years: A Practical Guide to Teaching and Learning*. London: Routledge, Taylor and Francis Group.
- HATCHER, P. J., HULME, C., MILES, J. M., CARROLL, J., HATCHER, J., GIBBS, S., SMITH, G., BOWYER-CRANE, C. and SNOWLING, M. (2006) Efficacy of small group reading intervention for beginning readers with reading delay: a randomised control trial. *Journal of Child Psychology and Psychiatry*, 47, 820–827.
- LLOYD, S. (1992) *The Phonics Handbook*. UK: Jolly Learning.
- SAVAGE, R. S., CARLESS, S. and STUART, M. (2003) The effects of rime- and phoneme-based teaching delivered by learning support assistants. *Journal of Research in Reading*, 26, 211–233.
- SAVAGE, R. S. and CARLESS, S. (2004) Predicting curriculum and test performance at age 7 years from pupil background, baseline skills and phonological awareness at age 5. *British Journal of Educational Psychology*, 74, 155–171.
- SAVAGE, R. S. and CARLESS, S. (2005) Learning support assistants can deliver effective reading interventions for 'at-risk' children. *Educational Researcher*, 47, 45–61.
- SAVAGE, R. S. and CARLESS, S. (2008) The impact of reading interventions delivered by teaching assistants on Key Stage 1 performance. *British Educational Research Journal*, 34, 363–385.
- TABACHNICK, B. G. and FIDELL, L. S. (2001) *Using Multivariate Statistics* (4th edn). New York: Allyn & Bacon.
- TORGENSEN, J. K. (2000) Individual differences in response to early interventions in reading: the lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, 15, 55–64.

Correspondence

Robert Savage
Faculty of Education
McGill University
3700 McTavish Street
Montreal
Canada
H3A 1Y2
Email: robert.savage@mcgill.ca

Copyright of Support for Learning is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.