# Benefits of training in linguistic awareness dissipate by grade 3?

PEKKA NIEMI
ELISA POSKIPARTA
MARJA VAURAS
*University of Turku, Finland*

ABSTRACT

The maintenance of phonemic awareness training effects was studied in a longitudinal setting extending over grades 1 through 3. A group of 11 readers-at-risk in grade 1 was superior to the controls, and comparable to average readers, in word recognition after training in grade 1, with the effect being sustained until the end of the grade (Poskiparta, Niemi, & Vauras, 1999). Towards the end of grade 2, the trained pupils were still superior to the controls in word recognition and lexical decision. However, no difference between these groups was observed in the speed and accuracy of reading aloud, with both groups performing more poorly than other preschool nonreaders who represented a reference group. By the end of grade 3, the differences between intervention and control groups had disappeared, with both groups performing more poorly than other preschool nonreaders even on lexical decision. The former also preferred watching TV over reading and read less versatile materials. Reading comprehension showed the superiority of the other preschool nonreaders both in grades 2 and 3. The results suggest that initial success in word attack skills does not promote useful reading practices in the absence of incentives. Therefore, the trained children slowly fall back to the level of their controls in measures tapping the automatization of reading. This constitutes a limitation to the bright prospects usually associated with the training of phonemic awareness.

*Key words:* Effect maintenance, Phonemic awareness, Training.

No doubt exists nowadays about the benefits of training in phonological awareness for subsequent reading acquisition. However, much less is known about the persistence of these effects. Do even the best interventions result only in temporary gains? Do pupils trained in preschool or in grade 1 still show superiority in decoding after a number of years of schooling? Is the gap to the average performance closing or widening? Is there any transfer to reading comprehension? Is it at all plausible that a one-shot training would place an intervention group permanently above controls, other things being equal? If not, when and how should phonological awareness training be followed by other forms of reading support? In other words, it would be informative to know what the initial gains are like after several years have passed since training. Is there a trend towards increasing differences between trained children and controls or, alternatively, do the gains disappear with time?

There is an enormous literature reporting beneficial effects of training. Some recent European research is described in a thematic issue (Niemi, Porpodas, & Tonnessen, 1999). A useful source for recent Anglo-American research is provided by Blachman (1997). Somewhat to our surprise, however, the vast

*Address:* Pekka Niemi, Department of Psychology, University of Turku, Lemminkaisenkatu 1, FIN-20014 Turku, Finland. E-mail: peknie@utu.fi

literature does not appear to cover the above considerations. This is aptly demonstrated by the comprehensive database of relevant European research that listed no less than 144 longitudinal studies since the 1980s (Schneider & Stengård, 1999). Of these, 44 included an intervention. However, the critical cross-section involving a controlled group intervention in phonological awareness and a subsequent follow up until the end of grade 2 was represented by only six studies. As could be expected, things got worse towards the end of grade 3, with only three relevant studies remaining in this category. The available evidence suggests that the word reading superiority of trained children can still be observed by the end of grade 2 (Borstroem & Elbro, 1997; Hurry & Sylva, 2000; Jimenez & Ortiz, in press; Lundberg, 1994; Olofsson, 1993; Petersen, in press; Schneider, Ennemoser, Roth, & Kuespert, 1999; Schneider, Kuespert, Roth, Visé, & Marx, 1997, Zorman, 1999). Four studies have followed up word reading ability as a function of phonological awareness training up to grade 3. The results suggest the usefulness of training even after such a long delay (Hurry & Sylva, 2000; Olofsson, 1993; Petersen, in press; Schneider, personal communication). Moreover, Kozminsky and Kozminsky (1995) reported sustained gains in reading comprehension, but their study did not include a measure of word reading ability. An encouraging feature is that in several of the above studies the intervention was focused on children-at-risk (Boerstrom & Elbro, 1997; Olofsson, 1993; Schneider et al., 1999).

The overall positive picture is somewhat blurred by the fact that measures of word reading were, without exception, based on word-to-picture matching. As a measure of decoding, this procedure certainly does a good job. However, it is a rather rough approximation of the process of word reading, giving clues from which weak readers might benefit, while good readers do not need them at all. The purpose of the present study was to test the notion that the success story may look different if latency measures tapping the automatization of reading are brought to bear on the measurement of reading competence. The study is an extension of a previous one that reported positive training effects for children-at-risk in grade 1 using both word-to-picture matching and latency variables (Poskiparta, Niemi, & Vauras, 1999). In that study, clear cut gains were observed for children-at-risk with a low verbal competence in a variety of tasks including word-to-picture matching, lexical decision, and spelling. The follow up based on a lexical decision task and the reading aloud of a text was continued until the spring of grade 3.

## Method

### Participants

The original subject pool comprised 240 children who entered the first grade in the autumn 1993, in four schools in Turku. Turku has 160,000 inhabitants and represents urban Finland well. Preschool readers were excluded from the present study, as were 10 children participating in an intervention in grade 2 and their respective controls. The intervention in grade 1 showed that about half of the participants did not really need it. They, and their respective controls receiving only classroom instruction, performed at an average level in criterion tests. These children were also omitted from the present study. There remained 11 children with a low cognitive profile (WISC-R Verbal Scale $z$ = -1.5 on the average). Their pairwise matched controls (preschool phonological awareness, listening comprehension and WISC-R Verbal Scale) comprised 11 pupils who received traditional special instruction, mainly revising the classroom instruction. Finally, 121 mainstream readers constituted two reference groups for comparisons. The inclusion criterion for the intervention group was a zero performance on a test of linguistic awareness four months prior to the start of schooling.

**Procedure**

In Finland, schooling starts at the age of seven. The typical teaching method for beginning readers is some variant of phonics, always emphasising the connection between sound and print. An approach like this is uniquely suited for the Finnish language that features an almost perfect grapheme-to-phoneme correspondence.

The intervention, essentially inspired by Lundberg, Frost, and Petersen (1988), was carried out during the autumn term of grade 1. During 13 weeks, the children participated in altogether 47 sessions of 20 min. duration. The program is reported in detail in Poskiparta et al. (1999). It consisted of four main components. These were clarification of self-image (Sessions 1-4), rhymes and nursery rhymes (Sessions 5-16), word and syllable awareness (Sessions 17-26), and phonemic awareness (Sessions 27-47).

**Assessment**

*Word recognition* was measured in grade 2 by means of the Finnish version of the Danish OS-400 test (see Lundberg et al., 1988). It includes 400 words written in a column and four pictorial alternatives for each of them. The reader has to choose the picture carrying the same meaning as the word. The session lasts 15 min. and the score is the sum of correct answers. The test was not given in grade 3 because of its simple nature, leading to doubtful face validity among advanced readers.

The *lexical decision* task required the child to decide, as quickly as possible, if a letter string on a computer screen is a word or a nonword. The targets in grade 1 were 30 four-letter words with two syllables. In grades 2 and 3, the target words consisted of 7 to 12 letters, the number of syllables varying from 3 to 5. Nonwords were formed by interchanging the order of syllables in real Finnish words.

*Story-reading speed and accuracy.* The child was asked to read aloud a 95-word narrative in the usual way. The instructions also emphasised comprehension. Two parallel stories were prepared for grades 2 and 3. The experimenter recorded the time the child spent on reading the story and the number of incorrectly read words. However, the word was scored as correct, if the child spontaneously corrected his or her misreading. The score was the average time spent on a word. The number of skipped words was subtracted from the total of 95. Reading accuracy was expressed by the number of incorrectly read words.

*Reading comprehension* was assessed in grades 2 and 3 with two parallel expository texts of 135 words. They described exotic hobbies, parachute jumping in grade 2, and diving in grade 3. The child was asked to read the text silently at his/her own pace so that s/he would be able to answer in writing five questions about its message afterwards. The answers were rated twice by two independent judges. The inter-rater agreement varied from .80 to .98, depending on the question. The discrepancies were resolved by discussion.

*Reading habits* were surveyed from grades 1 to 3 by simple questions inspired by Juel (1988). These focused on the child's view of the versatility of his/her reading materials, the choice between watching TV and reading, and the preference for reading over cleaning one's room.

**Results**

Grade 1 results for the Intervention and the Control group have been reported in Poskiparta et al. (1999). In short, clear-cut training benefits were observed in word recognition, as well as in lexical decision, both for words and nonwords. The performance of the Intervention group closely paralleled that of all other former preschool nonreaders. The question is whether these effects are sustained until the end of

grades 2 and 3. Tables 1 and 2 give the relevant data separately for grades 2 and 3. Because there was attrition among the control group by the third grade, the analyses are done separately for each grade.

**Grade 2**

*Word recognition* scores were subjected to an ANOVA which showed only a symptomatic interaction of Treatment x Reference Group, $F(1, 125) = 3.51$, $p = 0.063$. However, a planned contrast between the Intervention Group and their respective controls suggests a difference, $F(1, 125) = 4.23$, $p = 0.042$.

*Lexical decision* showed similar effects for words and nonwords, the latencies for the latter being, as expected, slower. Therefore, words and nonwords were pooled in the analysis. There was again only a symptomatic interaction of Treatment x Reference Group, $F(1, 129) = 3.44$, $p = 0.083$. The only group really differing from any other was the Control Group, with a planned contrast between it and the Intervention Group suggesting a stable effect, $F(1, 129) = 7.16$, $p = 0.008$. Errors of commission varied from 2% to 8%.

*Story-reading speed* failed to show a difference between the Intervention Group and the Control Group with the planned contrast being only symptomatic, $F(1, 131) = 2.96$, $p = 0.09$. All in all, the two reference groups were dramatically faster than the Intervention and Control Groups, $F(1, 131) = 30.40$, $p = 0.001$. The number of incorrectly read words showed an analogous pattern.

*Reading comprehension* effectively differentiated the Reference Groups from the Intervention and Control Groups, $F(1, 125) = 10.30$, $p = 0.002$. However, the planned contrast between the Intervention and Control Groups failed to reach significance.

**Grade 3**

*Lexical decision* showed predictable word superiority and a reliable difference in favour of the Reference Groups as opposed to the combined Intervention and Control Groups, $F(1, 124) = 19.35$, $p = 0.001$. This time, even the Intervention Group differed from its Reference Group resulting in a significant planned contrast, $F(1, 127) = 10.77$, $p = 0.001$. There was also a first-order interaction involving Word/Nonword and Reference/Intervention and Control Groups, $F(1, 124) = 8.25$, $p = 0.005$. This was due to a larger difference between words and nonwords for the Intervention and Control Groups than for the Reference Groups. The Intervention and Control Groups did not differ from each other in any comparison. Errors of commission varied from 2% to 6%.

*Story-reading speed* replicated the pattern obtained for grade 2. The Reference Groups as a whole again outperformed the Intervention and Control Groups, $F(1, 129) = 33.03$, $p = 0.001$. There were no other differences. Data on incorrectly read words showed a very similar pattern.

*Reading comprehension* again showed the superiority of the Reference Groups, $F(1, 114) = 19.96$, $p = 0.001$. There were no other differences.

*Reading habits* through grades 1 to 3 did not display differences between the Intervention Group and its respective controls. Therefore, the data were pooled for an analysis to try and determine whether there were any systematic differences between children-at-risk and mainstream readers on the whole. In grade 1, children-at-risk reported proportionally more often that they read only one type of text, usually cartoons, $x^2(2, N = 147) = 8.41$, $p = 0.015$. In grade 3, only one out of 18 children-at-risk reported reading expository texts such as factual books and encyclopedias as opposed to 19% of mainstream readers, $x^2(2, N = 134) = 11.11$, $p = 0.004$. Moreover, 89% of children-at-risk prefer-

**Table 1**
**Reading performance in grade 2**

| | Task | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Word recognition | | Lexical decision-word | | Lexical decision nonword | | Story-reading speed | | Story-reading errors | | Reading comprehension | |
| Group | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Intervention | 230.4 | 64.1 | 2677 | 873 | 4007 | 1238 | 1713 | 829 | 10.9 | 10.2 | 32.7 | 18.7 |
| Control | 169.0 | 78.3 | 4263 | 2438 | 6058 | 2994 | 2119 | 734 | 12.3 | 6.1 | 21.5 | 23.7 |
| Reference (intervention) | 240.0 | 60.2 | 2012 | 693 | 3046 | 1295 | 1160 | 460 | 4.8 | 4.7 | 43.1 | 17.4 |
| Reference (control) | 238.9 | 62.3 | 2421 | 1540 | 3675 | 2071 | 1285 | 458 | 4.7 | 3.8 | 41.8 | 19.3 |

*Note:* Word recognition max = 400; lexical decision = reaction time in milliseconds; story-reading speed = in milliseconds per word; story-reading errors = total number of errors; reading comprehension = percentage correct.

**Table 2**
**Reading performance in grade 3**

| | Task | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Lexical decision-word | | Lexical decision nonword | | Story-reading speed | | Story-reading errors | | Reading comprehension | |
| Group | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Intervention | 2017 | 776 | 3069 | 1289 | 1159 | 369 | 8.7 | 7.4 | 34.5 | 21.3 |
| Control | 2205 | 632 | 3551 | 1587 | 1320 | 320 | 8.6 | 8.5 | 31.8 | 9.7 |
| Reference (intervention) | 1417 | 442 | 2043 | 891 | 859 | 259 | 2.7 | 3.5 | 53.7 | 19.2 |
| Reference (control) | 1637 | 562 | 2432 | 988 | 892 | 224 | 3.7 | 3.1 | 52.4 | 15.6 |

*Note:* Lexical decision = reaction time in milliseconds; story-reading speed = in milliseconds per word; story-reading errors = total number of errors; reading comprehension = percentage correct.

red watching TV to reading in grade 3, as opposed to 57% of mainstream readers, $x^2(2, N = 129) = 6.74, p = 0.009$. Preference for reading over cleaning one's room increased from grade 1 to grade 3 with no significant difference between groups: children-at-risk from 43% to 71%, and mainstream readers from 61% to 82 %, respectively.

*Attrition due to school failure* occurred as an unintended result. Altogether three pupils from the original Control Group of 11 were either referred to a special school or repeated a grade after the first or second grade. The entire Intervention Group stayed with their classmates until the end of grade 3.

## Discussion

Although based on a successful intervention for phonological and phonemic awareness in grade 1, the present results are frustrating for the eager proponent of the phonological hypothesis. The gains obtained by children-at-risk in the autumn term of grade 1 were strong and well sustained until the end of grade 1 (Poskiparta et al., 1999). A year later the intervention children still outperformed the controls in lexical decision and word-to-picture comparison, and were also close to the performance of other preschool nonreaders in the latter task. However, reading aloud to comprehend a text placed both the intervention children and controls at the same level, and far behind the other preschool nonreaders in terms of speed and accuracy. What is more, even the difference between the intervention group and controls in lexical decision disappeared by the end of grade 3 with the intervention group now being inferior to its reference group. Reading aloud showed the earlier pattern, with children-at-risk performing somewhat more poorly than their mainstream classmates of the previous year.

The obvious difference between the present study and those showing training-related effects retained up to grade 3 is that the latter have relied exclusively on word-to-picture comparison. This task, while being easy to administer both individually and to groups, taps very basic word recognition ability, and even supports the latter by providing semantic cues. Matching pictures and words calls for a host of perceptual and motor processes, rendering the relative space of decoding processes small. Lexical decision leads to clearly faster response, indicating that the total latency includes a comparatively large word identification component. However, the fastest word identification times are observed while the child is reading aloud a continuous text. It is reasonable to surmise that among these three tasks, the level of reading automatization is effectively tapped only by reading aloud.

Hypothesising long-term transfer to reading ability from early training in phonological and phonemic awareness presupposes a huge leap from cracking the alphabetic code to the active use of the ensuing decoding skills. In fact, this would be equivalent to extensive reading practice. What would justify such a prognosis in the first place? Readers-at-risk tend to be verbally inferior to their classmates. In the present study their verbal IQ was $z = -1.5$, corresponding roughly to 7% from the lower end of the distribution. It appears unlikely, to say the least, that these children would start to read eagerly once they had mastered the code. In fact, our somewhat crude data on their reading habits point to the contrary. They tend to prefer simple reading materials and TV during their leisure time, thus differing from their mainstream classmates.

The level of reading accomplished by our children-at-risk was not bad in an absolute sense. The word-to-picture matching test allows a direct comparison with the Danish children in the study of Olofsson (1993). Scoring 169 correct in the spring term of grade 2, our readers-at-risk not receiving an intervention performed at the level of mainstream Danish readers some three months older. Their accuracy score in lexical

decision was better than that of representative Norwegian fourth-graders in Oftedal (2000). Their reading-aloud performance was reasonably fast and perfectly intelligible, errors notwithstanding. In other words, the question is about their relative standing in the class as slow but fairly accurate readers. In the course of three years of schooling, they show steady progress as readers but their initial success with word attack skills does not transfer to positive reading habits which constitute the basis for reading automatization. For this very reason, they gradually fall back to the level of the controls in measures of continuous reading. What is needed are longitudinal training studies of a very different character, namely, those tapping reading motivation. Phonemic awareness training helps the beginning reader to crack the alphabetic code. The resulting skill is not usually particularly fast. However, reasonable accuracy of performance is soon achieved, at least in the more regular languages. We suggest that this is about how far the beginning reader gets on the basis of an intervention in phonemic awareness. Starting to practise the art of reading is quite another matter related to the notion of transfer of training, scarcely touched upon in the phonemic training studies so far.

Do the present findings undermine the value of early phonemic training in an attempt to promote the reading acquisition of children-at-risk? Our answer is firmly negative. The mere fact that three control pupils out of 11 became dropouts supports this conclusion. Dropping out and being moved to another group puts stressful demands on the child. We did not measure the self-image of our pupils which, in retrospect, was an obvious omission. Passing grades at the same pace as classmates is a major accomplishment given an initial language inferiority. However, we suggest that with the proliferation of longitudinal studies featuring measures of reading automatization, the phonological training approach loses some of its former glory, though not all. Such a demonstration of the limits of a powerful approach should be most useful for paving the way for the vitally important study of transfer of phonological training.

## References


Blachman, B. (Ed.). (1997). *Foundations of reading acquisition and dyslexia. Implications for early intervention*. Mahwah, NJ: Erlbaum.

Borstroem, I., & Elbro, C. (1997). Prevention of dyslexia in kindergarten. Effects of phoneme awareness training with children of dyslexic parents. In C. Hulme & M. Snowling (Eds.), *Dyslexia: Biology, cognition and intervention* (pp. 235-253). London: Whurr.

Hurry, J., & Sylva, K. (2000). *Early reading intervention: Who does it help and for how long?* Paper submitted for publication.

Jimenez, J. E., & Ortiz, M. R. (in press). Metalinguistic awareness and reading acquisition in the Spanish language. *The Spanish Journal of Psychology*.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447.

Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. *Learning and Instruction, 5,* 187-201.

Lundberg, I. (1994). Reading difficulties can be predicted and prevented: A Scandinavian perspective on phonological awareness and reading. In C. Hulme & M. Snowling (Eds.), *Reading development and dyslexia* (pp. 180-199). London: Whurr.

Lundberg, I., Frost, J., & Petersen, O-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

Niemi, P., Porpodas, C., & Toennessen, F.-E. (Guest Eds.). (1999). Special Series: Prevention and treatment of dyslexia. *Journal of Learning Disabilities, 32*(5).

Oftedal, M. (2000). *Diagnostisering av ordavkodingsvansker: En prosessanalytisk tilnaermingsmåte* [Diagnosis of decoding difficulties: A process-analytical approach]. Unpublished doctoral dissertation, University of Bergen, Norway.

Olofsson, Å. (1993). The relevance of phonological awareness in learning to read: Scandinavian longitudinal and quasi-experimental studies. In R. Joshi & C. Leong (Eds.), *Reading disabilities: Diagnosis and component processes* (pp. 185-198). Dordrecht, The Netherlands: Kluwer.

Petersen, D. K. (in press). What do you get if you add /mmm/ to ice ? Training phoneme awareness in kindergarten. An intervention study with children of dyslexic parents. In N. Badian (Ed.), *Preschool prediction and prevention of reading failure*. Timonium, MD: York Press.

Poskiparta, E., Niemi, P., & Vauras, M. (1999). Who benefits from training in linguistic awareness in the first grade, and what components show training effects? *Journal of Learning Disabilities, 32*, 437-446, 546.

Schneider, W., Ennemoser, M., Roth, E., & Kuespert, P. (1999). Kindergraten prevention of dyslexia: Does training in phonological awareness work for everybody? *Journal of Learning Disabilities, 32*, 429-436.

Schneider, W., Kuespert, P., Roth, E., Visé, M., & Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. *Journal of Experimental Child Psychology, 66*, 311-340.

Schneider, W., & Stengård, C. (Eds.). (1999). Inventory of European Longitudinal Studies of Reading and Spelling. COST A8 - Learning disorders as a barrier to human development. Retrieved February 15, 2000 from the Worl Wide Web: http://www.psychologie.uni-wuerzburg.de/cost/inventory.htm

Zorman, M. (1999). Evaluation de la conscience phonologique et entravnement des capacites phonologiques en grande section de maternelle . [Assessment of phonological awareness and training of phonological abilities in kindergarten]. *Reeducation Orthophonique, 197*, 139-157.