This study was a quasi-experimental action research evaluation of a program for computerized self-assessment of reading comprehension (The Accelerated Reader). It looked at the formative effects on reading achievement and motivation in two schools in severely socio-economically disadvantaged areas. The results suggested that the program, even when less than fully implemented, yielded gains in reading achievement for these at-risk readers that were superior to gains from regular classroom teaching and an alternative intensive method, even with less time devoted to class silent reading practice than in comparison classes. Additionally, the program yielded significant improvement in measured attitudes to reading for girls.

Teachers are aware that assessment can lead instruction and learning as well as follow it—with good or bad results. Recent developments in computer assisted assessment of literacy seek to enable more frequent and detailed evaluations in less time with greater consistency, empowering the teacher and student by yielding more detailed formative feedback.

Many computer aided learning packages (e.g., integrated learning systems) include ongoing embedded curriculum based assessment. Free-standing, computer-based, norm-referenced tests of reading and baseline tests for measuring subsequent added value are paralleled by increasingly sophisticated diagnostic tests. Programs for recording, analyzing, and reporting on student progress are available, together with systems for electronic portfolio assessment and shell or authorware programs to enable teachers to create their own computer-based tests. Such programs are very different in their architecture within as

This research was supported by a grant from the Nuffield Foundation, London, England. Thanks are also due to the participating students, teachers, schools, and regional education authority in Scotland, and to G. A. Lindsay, Professor of Special Education at the University of Warwick (England), who was the external examiner for the project.

This paper is a development of a presentation given at the National Reading Research Center conference “Literacy and Technology for the 21st Century,” Atlanta, Georgia, 4 October 1996. The full research report lodged in ERIC (ED408567).

Address correspondence to Keith J. Topping. Centre for Paired Learning and Graduate Educational Psychology, University of Dundee, Dundee DD1 4HN, Scotland. http://www.dundee.ac.uk/psychology/kjtopping
well as between types, and they vary greatly in the extent to which they have been properly evaluated; as a result, broad generalizations about utility cannot be made.

The study reported below focused on another type of free standing computer assisted assessment, which was curriculum based, voluntarily self-administered by students, and specifically intended to have strong formative effects on subsequent learning.

THE PROGRAM

The study sought to evaluate the impact of the use of a program for computer assisted student self-assessment of practice at reading and ensuing comprehension. The software is known as The Accelerated Reader (AR; Advantage Learning Systems, 1993). It is designed to enable curriculum-based assessment of reading comprehension of real (“trade”) books by students and analyze and summarize the results. It aims to help teachers promote and manage effective reading practice.

Students using the program read a book from over 13,000 titles on the AR list and take a multiple-choice comprehension test on the book at the computer, which scores the test and keeps records. Each book has a maximum point value based on its length and difficulty. When the student self-tests, the computer awards points up to this maximum, according to their number of correct test responses. Students select their own preferred books and read at their own pace. Teachers may choose to allow students to test on books read to and with them, as well as those read independently and silently (which is especially useful in the case of early or delayed readers). The default condition is that students can test on a book only once. (Other software that might appear superficially similar generally has numerous differences in these and other organizational parameters, so practitioners should not assume equivalence.)

The quantitative feedback from the points system is intended to raise meta-cognitive awareness and motivate students to read more, longer, and harder books. The computer also provides the teacher with an automatically updated analysis of scores for individuals or whole classes, which indicates average percent correct, difficulty of books read, points earned, and other diagnostic information. Computer generated “at risk” reports, enable the teacher to guide each student’s reading practice for maximum effectiveness.

The software originated in the United States and is supported there by extensive staff development and in-service training opportunities, which lead to a wider educational development program
known as “Reading Renaissance.” AR is currently in over 40,000 schools in the United States, and its use is spreading to other countries. The associated “Model Classroom Program” identifying and celebrates classrooms in which good practice in the implementation of Reading Renaissance has been evidenced.

THEORETICAL UNDERPINNINGS

As indicated above, the program is intended to have an impact at two levels— that of the student and that of the teacher. AR is sometimes mistakenly assumed to be a “reading incentive” program, but the attachment of tangible reinforcement to points earned is not an integral part of the program, as this study demonstrates.

Considering the impact at a very simplistic level, the test does ensure that the student has actually read the book. More subtly, the points give the student fairly detailed and timely feedback on the success of self-chosen reading in terms of essential components understood. This enables greater student control over their reading activity in terms of management of appropriate challenge and other parameters of the reading process (Shapiro & Cole, 1994). Thus, meta-cognitive awareness is likely to be heightened, and feelings of self-efficacy as a learner are enhanced (Schunk, 1994; Schunk & Rice, 1993). AR can be accurately described as a “reading motivation” program, aiming to encourage students to read more, longer, and harder books through better quality and quantity of feedback.

For the teacher, detailed feedback on the reading performance of all students in the class is provided by the program with minimal effort. An immediate indication of the success of each student’s reading performance is available in relation to each student’s functional reading level and the number and difficulty of books they are choosing to read. Thus, the teacher is enabled to ensure that all students are reading within their own “zone of proximal development” (Vygotsky, 1978), and students who are currently reading ineffectively are quickly flagged as “at-risk” by the system. Of course, this high quality information on student learning is of no significance if it is not acted upon, and teachers are expected to intervene appropriately with students (through informal reading counseling or other guidance), subsequently using the system to track the effect of their intervention. AR thus constitutes an electronic “learning information system” that facilitates early teacher intervention and the evaluation of that intervention. It is intended to help teachers teach more effectively.
Feedback is clearly a key feature for both students and teachers, and Paul (1995, 1996) has elaborated this theoretical rationale in the context of cybernetic feedback systems. AR helps students operate, or be guided to operate, much more in their “zone of proximal development” (Vygotsky, 1978) as readers, maximizing challenge while minimizing failure and avoiding unproductive reading at levels too low or high for successful learning. While AR might increase levels of reading practice (see Topping & Paul, 1999), it seeks also to increase effective and accountable engagement with literature, rather than mere exposure to it. Readers are referred to these other works for a more detailed exposition.

PREVIOUS OUTCOME RESEARCH

A largely qualitative study was conducted by McKnight (1992), focusing on a population of at-risk readers and using multiple measures: observations, surveys, and questionnaires. The use of AR over a 12 week period was found to be associated with improved attitudes to reading in 17 fifth grade Chapter 1 students. However, many interventions other than AR were occurring in parallel, no data on AR implementation integrity were reported, there was no control or comparison group, and the possibility of Hawthorne effect is evident.

Large scale studies were conducted by Paul (1992, 1993). The first of these analyzed AR program and reading test data on 4,498 students aged 6—16 in 64 schools. The schools self-selected to respond to an invitation to submit their data for analysis (so their representativeness was questionable); reading test data were presumably on more than one test, and no information on implementation integrity is presented. Analysis indicated a strong positive relationship between the number of points accumulated through the AR program and gains in reading test scores, but, given the methodology, the direction of any causal linkage is equivocal. However, students with the lowest ability showed the greatest gains. A second study (Paul, 1993) extended this approach by including data from 10,124 students in first through ninth grade from 136 schools using twelve different standardised tests. Similar results were found for reading, and a positive relationship with increases in math scores was also evident. Younger and poorer readers appeared to improve more than older and more able readers.

Peak and Dewalt (1993) conducted a longitudinal evaluation over five years of the impact of AR with a relatively able group of 25 students enrolled up to ninth grade in one school. The program was only
available centrally in the school, and students could access it only out of class time. No process data were reported regarding implementation integrity. Using a matched comparison group of the same size from another school, reading test scores were compared when students were in third, sixth, and eighth grades. The AR group gained approximately twice as much as the comparison group and appeared to spend twice as much time reading, though exactly how the latter was measured is unclear.

A program to improve reading comprehension in 46 sixth through eighth grade underachieving students was developed and evaluated by Turner (1993). The program included AR, but the effect of this aspect was not partialled from other components. Pre-post testing of reading comprehension indicated 82% of participants had improved. Degree of increased reading activity was related to test gain. However, only eighth graders showed improvement in measured attitude to reading.

Longitudinal data on norm-referenced tests of reading and language administered state-wide were analyzed by Penuel (1997) for 19 elementary schools in a metropolitan area that reportedly had "used" AR for an average of almost two years. Actual gains were compared to expected gains. Students in both third and fourth grades made higher than expected gains in both language and reading, though these differences only reached statistical significance in language. Unfortunately, the unit of analysis appeared to be the school rather than the class or the student, and no information was presented on AR implementation format, range, or integrity. Additionally, the specific procedures used to determine statistical significance were not recorded. Penuel concludes that the longer the AR program had been in the school, the greater the gains in language, but the author reports only one significant correlation coefficient for support.

Other reports in the literature have been brief and anecdotal (e.g., Clements, 1995; DuBose, 1996; Lind, 1995; Smith, 1996).

OBJECTIVES OF THE PRESENT STUDY

The present authors were interested in investigating whether the AR software would prove effective in a different cultural setting, with at-risk readers. They also wished to carry out a somewhat more rigorous quasi-experimental evaluation of its effectiveness that tried to control for more of the potential confounding variables than previous studies (particularly, time on task at reading) and yielded stronger evidence about the nature and direction of causal linkages. Additionally, close scrutiny of the process of implementation was intended
to determine the extent that any effectiveness was a function of implementation integrity. However, the action orientation of the research subsequently necessitated some compromises.

**CONTEXT FOR THE PRESENT STUDY**

The program was introduced to two elementary schools in areas of low socio-economic status in Aberdeen, in north east Scotland. The schools were selected purposely on the basis of having the relevant hardware available in the classroom and being willing to participate. The classes involved included many children with reading delay, children with special educational needs, and those learning English as a second language.

All of the teachers were completely new to the program and received one day of initial in-service training and subsequent support and monitoring visits by a researcher. The pilot UK version of the software covered only 100 AR titles (what is considered “great literature” is culturally specific). The schools’ existing book stocks were audited to determine the presence of these titles, and some additional books were made available to the experimental classes to make up the shortfall. The AR books (those with tests available) were colour-coded to indicate their point value (length and difficulty).

Two separate quasi-experimental action research projects were undertaken, in schools A and B.

**PROJECT A**

**Context and Implementation**

Project A took place in one school in a mixed ability class of P7 (11-year-old, sixth grade) children. Comparison (“control”) children were drawn from a parallel mixed age class in the same school, of whom 12 were the same age as the experimental children. Both classes evidenced good ethos, had experienced teachers, and were considered by the principal to be comparable other than in age composition. The experimental class \( n = 27 \) had access to the AR on a Macintosh platform in class for 6 months from pre- to post-test, including access to AR books, public display in class of AR points gained by individuals (see Van Houten, Hill, & Parsons, 1975), and the opportunity to exchange points for tangible rewards (although the last of these was used very little, it appeared to do no harm; see Cameron & Pierce, 1994.) The quality of implementation of AR was moderately good after
a slow start. For the first 5 weeks, the AR class received on average only 15 minutes of reading time per day, whereafter this was increased to 30 minutes reading time plus 30 minutes of being read to by the teacher. Students could and did test on books read to them, with a high rate of success.

Insufficient lower readability AR books were available. The color-coding of book point value was not necessarily attended to by the children, nor was this initially encouraged by the teacher. Greatly enthused by the prospect of taking the computer test, initially some children selected books too hastily and read them too quickly, resulting in low test scores and rapid consumption of the limited book stocks available. Children were thus not necessarily reading within their “zone of proximal development.” AR “at-risk” reports were generated weekly but did not lead to systematic intervention by the class-teacher until after the first five weeks of the project. However, over the whole period, on average 84% of test items were answered correctly, compared to the AR designated optimum of 85–92%, so implementation drift may not have been too serious. Additionally, of 802 tests taken, 719 were passed, points gained totalled 1,961 (73%) out of 2,685 possible, and the average book reading level tested on by these sixth grade students was grade 4.1 (P5).

Two months into the project, training in cross-age peer-tutored “Duolog Reading” (a structured cross-ability technique for supported or assisted reading, Topping, 1995, 1997) was provided to the whole class, but this was subsequently used within class reading time for on average only two school days in every five, initially with 8 pairs but declining steadily to only two. The comparison children were all of the 12 P7 children in the parallel mixed-ability/mixed-age class. They were of very similar chronological age and gender balance to the experimental AR group, and had very similar pre-test reading ability on 2 out of 3 reading tests, though pre-test scores on 2 measures of attitude to reading were slightly less positive. They experienced regular classroom teaching of reading, which included 30 minutes of reading time per day throughout the experimental period. Also, the comparison students gave written feedback on book completion to the teacher on what they had read, and there was provision of extra special books for students of lower reading ability.

**Summative Evaluation**

Two norm-referenced reading tests were used to examine reading achievement gains that were well established in the U.K. and docu-
mented adequate reliability and validity in their respective manuals. Neither had been in use in the experimental schools before. Pre-test, post-test, and follow-up scores for experimental and comparison groups are given in Table 1. On the Shortened Edinburgh group reading test of silent reading comprehension (University of Edinburgh, 1985), standardized scores (reading quotients) for the AR group showed a statistically significant increase over the experimental period, while those of the comparison group did not (non-parametric tests, two-tailed, throughout). Inspection of the gains makes clear that this is not attributable to differences in sample size. On parallel forms of the Neale individually administered test of oral reading accuracy and reading comprehension (Neale, 1989), a random sample of 12 AR children showed a statistically significant increase in reading accuracy over the experimental period, while the comparison group did not. On the comprehension scale of the Neale test, the AR sample group did not show a statistically significant change, but the comparison group showed a statistically significant decrease.

<table>
<thead>
<tr>
<th>TABLE 1 Project A Mean Pre-test, Post-test, and Follow-up Scores on Edinburgh and Neale Reading Tests for Experimental and Comparison Groups.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edinburgh (standardized score)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
</tr>
<tr>
<td>Post-test</td>
</tr>
<tr>
<td><strong>Neale Accuracy (reading age, months)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
</tr>
<tr>
<td>Post-test</td>
</tr>
<tr>
<td><strong>Neale Comprehension (reading age, months)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
</tr>
<tr>
<td>Post-test</td>
</tr>
</tbody>
</table>

*Note.* Standard deviation in ( ); n in [ ].
Although experimental and comparison groups had not differed significantly in Neale comprehension at pre-test, they did show a statistically significant difference at post-test.

Although, the AR group had clearly performed better than the comparison group on the comprehension scale of the Neale test the results raised doubts about the comparability of the supposedly parallel forms. A further follow-up Neale test was therefore undertaken 3 months after the post-test, using the first of the parallel forms again. Use of AR had continued during the follow-up period. At follow-up both groups showed gains in comprehension, but the experimental group remained significantly superior to the comparison group.

Two measures of attitude to reading were utilized: the Elementary Reading Attitude Survey (ERAS; McKenna & Kear, 1990) and a locally devised Reading Interest Scale. Overall, the AR group showed greater improvement in attitude to reading than the comparison group on both measures, but these differences did not reach statistical significance. However, gender differences were apparent. At pre-test, AR girls showed statistically significantly better attitudes to reading than boys on ERAS. At post-test, AR girls performed statistically significantly better than boys on both measures, demonstrating even bigger differences. (Virtually no gender differences were found in reading test scores.)

PROJECT B

Context and Implementation

Project B took place in another elementary school, in which no parallel class including same age students, was available. Hence, the experimental class contained mixed ability P7 (11 year old, sixth grade) children and the comparison class, which received an alternative treatment, contained P6 (10 year old, fifth grade) children. The experimental AR class evidenced a poor ethos, while the comparison class evidenced a good ethos. Both classes had experienced teachers.

The AR class \(n = 24\) included a preponderance of males (14:10). They had access to the AR on an IBM platform in class for 6 months from pre- to post-test, including access to AR books and public display of AR points, but there was no possibility of exchanging these for tangible rewards. The quality of implementation of AR was adequate only towards the end of the project. For the first half of the project period, the students had only 15 minutes of reading time four times per week, rising to 20–30 minutes for the last three months.
As in Project A, insufficient lower readability AR books were available, and the color-coding of book point value was not attended to by the children or initially encouraged by the teacher. Some children selected books too hastily and read them too quickly, resulting in low test scores and rapid consumption of AR book stocks. Children were thus not necessarily reading within their “zone of proximal development.” AR “at-risk” reports were generated weekly but did not lead to any intervention by the teacher until the second half of the project and systematic intervention only at the end of the project. However, over the whole period, on average 84% of test items were answered correctly, compared to the AR optimum of 85–92%. Of 701 tests taken, 610 were passed, points gained totalled 1,093 (60%) out of 1,833 possible, and the average book reading level tested on by these sixth grade students was grade 3.4 (P4.4).

At pre-test, the younger alternative treatment class \((n = 26)\) were found to be on average much more able readers relative to their age than the AR class. On the Neale comprehension pre-test, the younger class actually had a much higher mean reading age than the experimental class, despite the age difference. The alternative treatment (AT) class also included a preponderance of females \((16:10)\). On one reading attitude measure (ERAS), the AT class scored much higher at pre-test than the AR class, but on another, they scored slightly lower.

The alternative treatment consisted of regular classroom teaching supplemented by 15 minutes of individual silent reading time per day and group oral reading on a restricted selection of novels. The latter involved a low reading ability group in the class reading aloud to each other for 20 minutes three times per week, a middle ability group for 20 minutes twice per week, and a high ability group for 20 minutes once per week. During the group reading sessions, the classroom teacher visited with groups and quizzed them on the content of their reading. There was a public display of books read by students. Additionally, homework assignments were given at the end of each book chapter, involving written comprehension tests. The teacher also created reading comprehension puzzles, worksheets, secret messages, and so on, which were completed in class or as homework.

**Summative Evaluation**

On the Shortened Edinburgh group reading test of silent reading comprehension, standardized scores (reading quotients) for both the AR and the AT group showed a statistically significant increase over the experimental period, and the AT group shower a larger gain from a higher baseline (see Table 2).
### TABLE 2 Project B Mean Pre-test, Post-test and Follow-up Scores on Edinburgh and Neale Reading Tests for Experimental and Alternative Treatment Groups.

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Alternative group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Edinburgh (standardized score)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>89.46 (19.06) [24]</td>
<td>93.69 (13.77) [26]</td>
</tr>
<tr>
<td>Post-test</td>
<td>92.59 (15.78) [22]</td>
<td>99.96 (15.00) [26]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Alternative group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neale Accuracy (reading age, months)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>109.36 (24.20) [11]</td>
<td>106.67 (29.73) [12]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Alternative group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neale Comprehension (reading age, months)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>96.45 (20.72) [11]</td>
<td>103.25 (21.87) [12]</td>
</tr>
</tbody>
</table>

*Note.* Standard deviation in ( ); n in [ ].

On parallel forms of the Neale individually administered test of oral reading accuracy and reading comprehension, random samples of 12 AR and 12 AT children showed no statistically significant increases in reading accuracy over the experimental period, though the gain of the AR group was higher from a slightly higher baseline. On the comprehension scale of the Neale test, the AR group showed a very large statistically significant increase from a lower baseline, while the AT group showed a small decrease from a much higher baseline that did not reach statistical significance.

On the Elementary Reading Attitude Survey, the AR group showed a greater improvement in attitude to reading than the AT group, from a lower baseline. On the Reading Interest Scale, however, the AT group showed a much greater improvement in attitude to reading than the AR group (which changed very little), from a slightly higher baseline. None of these differences reached statistical significance. However, statistically significant gender differences were apparent, with AR girls showing better reading attitudes on both measures than boys at post-test though not at pre-test. Alternative treatment (AT) girls showed significantly better reading attitudes on one
measure (ERAS) than boys at post-test, though not at pre-test. Virtually no gender differences were found in reading test scores.

SUMMARY AND DISCUSSION

Clearly, this study can at best be described as quasi-experimental since it was very imperfectly controlled, though in this action research context, this was itself beyond the control of the researchers. In particular, the classes in the two projects were not wholly comparable. They appeared so in project A, at least so far as the pre-test ages and scores, classroom observation of ethos by the researchers, and the perceptions of the principal were concerned, but it is possible that other confounding variables might have intervened. In project B, the classes were different in age (though much less so on pre-test scores) and were very different in ethos and gender balance, although this seems likely to have mitigated against finding experimental effects. Also, the classes were actively given different new interventions by different teachers. Experimental teachers themselves had special treatment by way of training and follow-up, while comparison and alternative treatment teachers did not. The non-experimental classes had equivalent access to computers (but not AR software), but the time spent on-task at the computer was not controlled. Despite the acknowledged possibility of many other variables intervening, it is argued that conducting the two projects in parallel does strengthen the overall findings.

The two experimental AR classes also experienced several difficulties: the very restricted nature of the pilot UK version of the program, the insufficiency of books, the brevity of the initial training received by the teachers, and that the program was not properly implemented. Nevertheless, the results were encouraging.

In project A, despite having less class reading time than comparison students, the AR group showed significant gains on one test of reading comprehension and one of reading accuracy, while comparison students did not; another test of reading comprehension showed gains significantly higher than those of comparison students. AR children showed greater improvement in reading attitude than comparisons, significant only for girls. The minimal use of extrinsic reward and Duolog Reading was not thought to have significantly affected these results.

Results from project B were more complex. Comparison between the AR class and the alternative treatment (AT) class (which was younger, much more able in reading relative to age, much more pre-
ponderantly female, and had a better ethos) is indeed problematic. However, on two norm-referenced tests of reading comprehension, the AR group showed significant gains on both, while the AT group showed gains on only one, despite the fact that the AR group had less class reading time than the AT group (and less than the AR group in project A). Surprisingly, neither group showed significant gains in reading accuracy. Reading attitude results at pre-test and post-test were conflicting, and gains were significant only for girls. The alternative treatment was very labor intensive for the teacher but did not produce as good reading test results as AR, and its instructional efficiency is therefore questionable.

CONCLUSIONS

Taking both projects together, the results suggested that the Accelerated Reader program, even when less than fully implemented, yielded gains in reading achievement superior in a number of respects to both regular classroom teaching and an alternative intensive method, even with less time devoted to class silent reading practice than in comparison classes. Additionally, the program yielded significant measurable gains in attitudes to reading for girls.

The current study thus suggests that the AR is effective by means other than merely increasing the quantity of reading practice (time on task at reading). Its aim is to do this by also improving the quality of engagement with literature by students. Additionally, the study suggests that the AR program can be effective regardless of the availability of extrinsic tangible reinforcement, which elicited virtually no interest in the socio-economically disadvantaged Project A students despite the best efforts of an enthusiastic teacher (though this finding might be locally and/or culturally specific).

FUTURE RESEARCH

Further replication is clearly required in a variety of other contexts and in better controlled studies. Studies of the effectiveness of the Accelerated Reader might focus on experimental variation in the quantity of class reading time and consequent reading achievement. Variation in the quality of implementation likewise merits further study, particularly with respect to actual use of color-coding of book point value and active teacher response to “at risk” reports. The relative effectiveness of an expanded UK version involving availability of more books and tests also requires exploration.
Additionally, the additive value of including Duolog Reading in class reading time on a substantive basis requires experimental investigation, as does the effect of other forms of reading to and with students, including electronic forms. Beyond this, experimentation with alternative summative measures of reading comprehension and attitude might be considered. Comparison of effectiveness and instructional efficiency with other alternative treatments would also be of interest. Further inquiry into gender differences in effectiveness and the impact of public display (versus private self-recording) of AR performance is also needed, though findings may well be culture-specific.

REFERENCES


