A Review of the Literature on Computer-Assisted Learning, Particularly Integrated Learning Systems, and Outcomes with Respect to Literacy and Numeracy

Report to the Ministry of Education
Auckland UniServices Ltd
A REVIEW OF THE LITERATURE ON COMPUTER-ASSISTED LEARNING, PARTICULARLY INTEGRATED LEARNING SYSTEMS, AND OUTCOMES WITH RESPECT TO LITERACY AND NUMERACY

FINAL REPORT

AUCKLAND UNISERVICES LIMITED
a wholly owned company of
THE UNIVERSITY OF AUCKLAND

Prepared for:
Ministry of Education
45-47 Pipitea Street
Level 4
Thorndon
Wellington

Date: November 2000

Prepared by:
Dr Judy M. Parr
Irene Fung
School of Education
University of Auckland
# Table of Contents

Table of Contents....................................................................................................................iii
List of Tables...........................................................................................................................iv
Executive Summary....................................................................................................................v

## Part A: Computer-Assisted Learning and Learning Outcomes ............................................... 1
  1.0 Methodological Issues ................................................................................................ 1
  1.1 Approach taken in this report ................................................................................... 1
  1.2 Terminology .............................................................................................................. 1
  1.3 Design and CAL/ CAL / ILS research ........................................................................ 3
  1.4 Evaluating across studies: The use of effect size..................................................... 4
  2.0 CAL and Learning Outcomes ...................................................................................... 5
    2.1 Introduction............................................................................................................... 5
    2.2 Larger scale analyses and meta-analyses ............................................................... 6
    2.3 Putting a face to meta-analyses: Specific reading and maths CAL software ......... 14
    2.4 Longitudinal and larger scale studies of CAL ......................................................... 17
    2.5 The special case of Integrated Learning Systems (ILS)......................................... 19
    2.6 Ability and ILS findings ........................................................................................... 33
    2.7 Summary of ILS ...................................................................................................... 35
    2.8 Conclusions and discussion ................................................................................... 36

## Part B: The Value of CAL / ILS ............................................................................................. 38
(Relative to Other Computer and Non Computer Innovations) ............................................. 38
  1.0 Introduction............................................................................................................. 38
  1.1 Relative effectiveness ............................................................................................ 38
  1.2 Relative costs ......................................................................................................... 38
  1.3 Alternative programs and costs.............................................................................. 40
  1.4 Non-cost, non-effectiveness related value of ILS................................................... 42

## Part C: Preconditions for Effective Use................................................................................. 44
  1.0 Introduction................................................................................................................ 44
    1.1 Organisational aspects ........................................................................................... 44
    1.2 Pedagogical ............................................................................................................ 46
  2.0 The New Zealand Experience: Views of Schools...................................................... 51
    2.1 Context ................................................................................................................... 51
    2.2 Participants and procedure .................................................................................... 51
    2.3 Issues from the discussions that suggest preconditions likely to be associated with
        more effective outcomes ........................................................................................ 52
    2.4 Discussion .............................................................................................................. 59

## Part D: Reflection and Recommendations............................................................................ 61
  1.0 Recommendations..................................................................................................... 63

References............................................................................................................................ 64

Appendices ........................................................................................................................... 73
  Appendix 1......................................................................................................................... 73
  Appendix 2......................................................................................................................... 74
  Appendix 3......................................................................................................................... 81
  Appendix 4......................................................................................................................... 85
List of Tables

Table 1
Findings from Eleven Meta-Analyses on Computer-Assisted Instruction (1978-1985)......................... 8

Table 2
Findings from Twelve Meta-Analyses on Computer-Based Instruction ................................................ 9

Table 3
Mean Effect Sizes and Percentile Gains for Nine Study Feature Variables and Other Variables ....... 11

Table 4
Mean Effect Size in Subject Areas .......................................................................................................12

Table 5
Mean Effect Size by Differing Years....................................................................................................13

Table 6
Top Ten Mathematics CAI Studies .......................................................................................................16

Table 7
Summary of Sample Profile and Effect Sizes for Mathematics: Phases 1 - 3a .....................................31
Executive Summary

Overall, the effectiveness of computer-assisted learning has not been conclusively demonstrated. To date, it has been shown to be less effective, on average, than other forms of intervention in education. In considering the results of evaluative research in computer assisted learning, one has to avoid confounding the medium with the method. Generally, computer-assisted learning software is underpinned by an older, neo-behaviourist theory of learning, one that has been displaced in the classroom by more social constructivist views of learning. Particularly in New Zealand primary classrooms, the approach of the software may differ considerably from widely accepted classroom pedagogy.

Computer-assisted learning programs, especially integrated learning systems, are generally costly. Their efficacy and cost effectiveness relative to alternative programs, particularly with respect to reading, is questionable. While comparative research exists with respect to effectiveness, good comparative research in relation to cost effectiveness is lacking.

Results from evaluations of integrated learning systems show highly variable results, with independent evaluations tending to be less favourable. The best results appear to be for basic maths skills; there is little evidence of gains in reading. Integrated learning systems, in their current form of neo-behaviourist, mastery learning, support the gaining of basic procedural knowledge. There is evidence that students may not be able to apply such knowledge without teacher intervention and that such knowledge may not generalise to school or system curriculum assessment tasks.

Part of the variability in outcome results stems from the different off-system assessment measures used to measure progress and part stems from the differing contexts of implementation. The latter includes characteristics of the student body and organisation for implementation including configuration of resources and deployment of personnel. Above all, this latter factor concerns integration, particularly the match between CAI or the integrated learning system curriculum content and methods, and that of the school and classroom.
Research questions

This research sought to answer the following:

1. What does the literature say about the effects of computer-assisted learning on student outcomes in literacy and numeracy?
2. What does the literature say the effect is relative to other types of interventions?
3. What does the literature say about preconditions for computer-assisted learning to be optimally effective in relation to student learning outcomes in literacy and numeracy?
4. What are school that have experienced computer-assisted learning reporting about their views of the pre-conditions necessary for effective use of computer-assisted learning again in relation to student learning outcomes in literacy and numeracy?

Overview

Part A begins with a description of the approach taken to answer the research questions. This is followed by a discussion of issues relevant to a consideration of computer-assisted learning (CAL) and, in particular, its effects on learning outcomes. First, there is a discussion of definitions of terms like computer-assisted instruction (CAI), computer-assisted learning (CAL), computer-based learning (CBL) and integrated learning system (ILS). In light of this discussion, the scope of the review is clarified. Other issues relate to the extent and veracity of the available evidence. Then, the aim is to assess the evidence regarding the use of computer-assisted learning in literacy and numeracy. The major findings from the use of CAI, largely from studies employing meta-analyses are discussed. Next, to put a face to the meta-analysis research, a selection of studies of effectiveness of widely used products is presented. Studies of a longitudinal and extensive nature conclude this section. This general discussion of computer-assisted learning is followed by a more detailed consideration of integrated learning systems (ILS), which are a major focus of this report.

The second part of the report considers the relative effectiveness of CAL. At its most basic, this includes evidence from research studies that have attempted, in some way, to assess the value of CAL. Any study where cost and effectiveness are considered and any research relating the use of computers to other innovations likely to be employed in addressing the enhancement of learning outcomes in literacy and numeracy is discussed. In particular, data relating to the relative effectiveness of ILS to other innovations are presented.

The final part of the report is important in that it examines the preconditions for effective use of CAL in general and ILS in particular. This part is premised on the notion that
even if CAL were designed to operate alone, using technology to improve learning is not just a matter of examining hardware and software features. Technology is a cultural object that functions in a social context where teacher and peer influences also operate. Thus this part aims to delineate the nature of influences, both positive and negative, that operate. The final section in this part discusses the data gathered from interviews with New Zealand schools regarding their views of conditions conducive to obtaining maximum benefit from an ILS.
Part A: Computer-Assisted Learning and Learning Outcomes

1.0 Methodological Issues

1.1 Approach taken in this report

The aim in this project was to locate recent information relevant to the research questions posed and to ensure that the material was both representative and of quality. Recent material is an important consideration when dealing with a moving target such as technology. To ensure that the literature was representative of the state of knowledge in the area, computer-based searches were conducted using the major databases in the social sciences (e.g., ERIC, PSYCLIT). As well, recent material was obtained from a number of web sites of organisations like the Educational Testing Service and the Milken Family Foundation. The preference was for published studies as a quality control measure.

Our computer-based search and selection of the literature was limited to those studies that are related to using computer technology to deliver curricula from kindergarten to year 13 with the emphasis on years 1-8.

The volume of research literature available, while considerable and growing, does not compare to that available in a field like reading. Relative to the literature available on reading, for example, the literature concerning technology and literacy has been scarce. In a recent overview, Kamil and Intrator (1988) estimated the amount of literacy research considering technology to be between one and five percent.

The nature of the literature is another important consideration. Much of what has been written on computers in education is not strictly research or scholarship but rather comment or the reporting of informal observations and opinions. Leading scholars point out that there is vastly more intuitive speculation in pieces that resemble position papers than there is systematic research in an area like literacy and technology (Leu & Reinking, 1996).

1.2 Terminology

In the research literature, computer-assisted instruction (CAI) is a generic term that includes a range of forms, varying according to different implementations of computer technology to assist instruction. Computer-assisted learning effectiveness has also been variously defined. Definitions of effectiveness may include academic (e.g. achievement) or non-academic (e.g. learning motivation) outcomes, or efficiency (e.g. reduced learning time),
or cost-effectiveness. Such being the case, embedded in the terminology of CAI are, in fact, variously defined constructs. The focus of this report is on the effectiveness of CAI in literacy and numeracy learning outcomes.

Since the advent of microcomputers and instructional software for education, computer-assisted instruction (CAI), computer-assisted learning (CAL), or computer-based instruction (CBI) has provided a supplemental instructional method in schools. There are not precise definitions of the terms CAI, CAL, and CBI. Generally, the concept of CAI in the early research was aligned to “drill-and practice” programs (Cognition and Technology Group at Vanderbilt, 1996). CAL includes more sophisticated programs which incorporate tutorial instruction (Wright & Marsh II, 1999-2000). Many CAL systems also include record keeping and management systems. However, CAL also goes by a variety of other names, such as CAI and CBI. With respect to CBI, it places emphasis more on individualisation of the learning process to accommodate the needs, interests, proclivities, current knowledge, and learning styles of the students. CBI software consists of tutorial, drill and practice and, more recently, Integrated Learning Systems (John Schacter, 1999). However, given the rapid development of technology, recent types of courseware were not available in the early CAI research. Modern implementation of CAI includes more advanced hardware and software technology, and allows for greater student interaction, and greater stores of information (Christmann, Badgett, & Lucking, 1997). In the more recent evaluations of research on computer-assisted learning, CAI is more a generic term covering drill-and-practice, tutorials, simulation/interactive thinking, word processing, conferencing, and other activities (Fletcher-Flinn & Gravatt, 1995).

An integrated learning system (ILS) is a far more specific term and such software can be differentiated from other computer-delivered curriculum content in that it has additional components, namely an extensive record and management system. However, more recently, an ILS is seen as having the ability to present learners with a question, process the answer and then give feedback before selecting an appropriate subsequent question.

An ILS is a computer-based system that manages the delivery of curriculum materials to pupils so that they are presented with individual programs of work over a number of weeks and months... It has three main components to facilitate the management of learning by teachers:

**Curriculum content**: This comprises an extensive range of tutorial, practice and assessment modules for a substantial part of a pupil's curriculum with coverage across a range of curriculum subjects and levels of ability.
A pupil record system: This maintains information on every pupil and records pupil's levels of achievement.

A management system: This links and controls the flow of data and may perform some or all of the following functions:
1. interpretation of pupil responses in relation to the current task;
2. updating of pupil records;
3. choice of pathways through the curriculum content;
4. delivery of the appropriate sequence of learning modules;
5. provision of feedback to pupils and teachers (Brown, 1997, p.7).

1.3 Design and CAI/ CAL / ILS research

Research on the impact of technology on learning is still relatively recent but, more so than with respect to other innovations, we are researching a moving target given both the pace of technological change and the fact that the introduction of technology to an educational setting is an ongoing, evolving process. There have also been problems with evaluations of technology, including the fact that it has been often treated as an undifferentiated variable and, further, as an independent variable. Evaluations have often suffered from poor design and poor measures of outcomes. But, percolating through has been a view that the features of the software; the features of the context, including the students, the teacher and the way technology is used, and the interactions among these are the vital considerations.

Although findings from meta-analyses and other studies comparing CAI with “traditional instruction” generally show a small advantage in favour of CAI over traditional instruction, as research on CAI effectiveness progressed, it became imperative for researchers to find answers to the question of why any CAI advantage occurred. Findings from meta-analyses that attribute effect to the medium (computer) alone have ignited considerable debate among researchers. Clark (1983), for example, believes that instructional methods embedded in the medium influence learning. To support his argument he found consistent evidence from his review of previous meta-analyses and other studies of media's influence on learning. He explained that studies comparing CAI with traditional classroom instruction were basically meaningless because they hopelessly confounded media and instructional method. “If media were conducted under rigorous controls,” he speculated, “the method and not the media would prove consequential.” Clark (1983) concluded that “the media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in
our nutrition” (p. 445). The corollary of this argument is that the relative merits of employing media in education are only economic factors associated with access and speed of delivery rather than pedagogical or cognitive benefits.

On the other hand, Kozma (1991) believes that the medium and methods combine to interact with and influence how students learn and process information. He suggests that future studies attempt to understand the media’s relationship to learning rather than its effects on learning. He reasons that learning with media can be thought of as “a complementary process within which representations are constructed and procedures performed, sometimes by learner and sometimes by the medium” (Kozma, 1994, p. 11).

These controversies lead Means and colleagues to write “The accumulation of comparative studies, biased on their choice of control groups or outcome measures, does little to help us understand what features of the treatment are critical for producing the desired effects (Means et al., 1993, p. 76). These limitations have given rise to longitudinal studies and other contextual approaches which demonstrate technology’s potential for supporting a learner-centred approach to education (Means et al., 1993).

1.4 Evaluating across studies: The use of effect size

An issue in consulting a large literature is the varying ways in which the research reports the outcomes of computer-assisted learning. In order to attempt to make sense from a diversity of types of statistics contained in different reports, a measure is often used by researchers that allows different types of information about effects to be converted into a common measure of effectiveness, namely, an effect size. The use of effect size allows us to combine results from multiple similar individual research studies (in this case concerning the use of computer-assisted learning) to generate a single effect size that illustrates the treatment effect across all studies. It provides a common expression of the magnitude of study outcomes for all types of outcome variables that, in this case, are achievement outcomes. It also allows a comparison of quite different types of innovations in learning, like the use of reinforcement; parent involvement in the school and the use of peers in learning.

Effect size employs a continuum where zero means that there is no effect from introducing some innovation, while a negative effect indicates that the innovation has a decreased effect on achievement and a positive effect that an innovation has an increased effect on achievement. It is calculated to determine the presence of a statistical difference in mean standard deviation units. There are many possible ways to estimate effect sizes (for a
The following is one of the fundamental formulae to calculate effect sizes:

$$E_s = \frac{M_e - M_c}{S_c}$$

In this formula, $M$ is the mean and $S$ is the standard deviation. An effect size of 1.0 indicates an increase of one standard deviation. Effect sizes are often expressed as percentiles or percentage improvement in learning to help with interpretation of what a particular effect size means. For example, an effect size of 1.0 indicates that 84% of the treatment group performed better than subjects in the control condition who scored at the mean.

There is the question of what is a reasonable effect size in terms of classroom significance. This is a difficult one. Hattie (1990, 1992, 1999) reported a synthesis of 337 meta-analyses, 200,000 effect sizes from 180,000 studies representing more than 50 million students and covering almost all types of innovation in education. His conclusion was that most innovations that are introduced to schools improve achievement by, on average .4 of a standard deviation. This provides a benchmark figure by which to judge effects as it is based on the effects of actual educational innovations. Other researchers have different ideas. Cohen (1977) provided the following ranges for mean effect size interpretation: effect size (ES) .2 to .49 = small effect; ES .5 to .79 = medium effect, and ES .80 and above = large effect. Tallmadge (1977) suggests that an effect size of .25 or more is educationally significant.

2.0 CAL and Learning Outcomes

2.1 Introduction

In this section on CAL and learning outcomes in literacy and numeracy we concentrate on recent major research output, either meta-analyses or large scale comprehensive, and likely to be generalisable, studies of CAI or studies of widely used software. A separate section is devoted to ILS and, given its primacy in terms of this report, smaller scale studies are included.

There is a general belief that computer technology allows educators many options for communicating, facilitating, and enhancing teaching and learning. Despite the rapid development of computer technologies and the increasing use of computer technologies to
deliver instruction, the debate about whether or not media enhance learning has continued unabated for the last 15 to 20 years. The central debate issue is the importance of delivery systems and instructional methods. Proponents claim that computer technology makes learning easier, more efficient, and more motivating (Schacter & Fagnano, 1999). They draw support from research reports and reviews that show a positive learning advantage for computer-assisted instruction when compared with traditional instruction (Fletcher-Flinn & Gravatt, 1995). Sceptics argue that if media studies were conducted under rigorous controls, the ‘instruction method’ and not the ‘media for instruction’ would prove consequential (e.g., Clark, 1983, 1994; Neal, 1998).

The challenge for educators is that studying the relationship between computer technology assisted instruction and learning outcomes has been complicated by changing emphases resulted from the development of learning theories, teaching methods, and technologies, as well as their differential impact on individual students. However, to produce meaningful learning, it is necessary to have a clear understanding of which technologies under what conditions are best suited to enhance learning and achievement of which group of students.

2.2 Larger scale analyses and meta-analyses

There were several large-scale evaluations of CAI conducted in the late 1960s and 1970s (e.g., Alderman, 1978; Atkinson, 1968; Murphy & Rhea-Appel, 1977; Suppes & Morningstar, 1968). Findings from these early studies, however, were inconsistent and clear conclusions were not often provided. For integrating results, early reviews used a box-score technique and reported the proportion of studies that were favourable and unfavourable toward CAI, as well as descriptive comments on the studies. For example, Vinsonhaler and Bass (1972) summarised the results of 10 major studies conducted between 1967 and 1970 involving CAI drill-and-practice with over 10,000 elementary students. They concluded that students who received CAI drill-and-practice generally demonstrated gains of 1 to 8 months over students in control groups who received traditional instruction. Their conclusions were supported by a later box-score review conducted by Edwards, Norton, Taylor, Weiss, and Dusseldorp (1975) who assessed the effectiveness of drill-and-practice, problem-solving, simulation, and tutorial CAI programs in student achievement. Based on six studies, they concluded that CAI plus traditional instruction was more effective than traditional instruction alone. Jamison, Suppes, and Wells (1974) also report similar results in their box-score review. They concluded that students at elementary level, disadvantaged students in particular, benefited when CAI was used as a supplement to traditional instruction. At the
secondary and college level, CAI was at least as effective as traditional instruction, and in some cases CAI resulted in substantial savings in student time.

Although box-score reviews reported ‘how often’ a particular method came out on top, they did not report ‘how much better’ one method was than another. Nor did they use statistics to find the characteristics that distinguished studies with positive results from those with negative findings. To address these limitations, a more sophisticated meta-analysis approach was introduced to CAI research, where differences between treatments were reported in effect sizes (Cognition and Technology Group at Vanderbilt, 1996).

Meta-analysis as a technique is an integrative statistical analysis or reanalysis of previous research as a means to answer new questions using old data (Glass et al., 1981). As discussed above, it combines results from multiple similar individual research studies to generate a single effect size that illustrates the treatment effect across all studies, so as to make studies comparable.

A number of meta-analyses have been conducted since this method was introduced to CAI research. However, effect sizes reported are quite variable. A possible reason is that meta-analysis can have multiple representation of a study in a data set. If studies report more than one finding as the result of using several experimental or control groups, or several tests or subscales, sample sizes will tend to be inflated and less reliable (Kulik & Kulik, 1991). Moreover, there has been a concern that the data used to calculate effect sizes in many meta-analyses might come from studies that are methodologically flawed.

There are several recent major reviews of meta-analytic reviews that may provide an updated overview of CAI effectiveness since 1970s (e.g., Fletcher-Flinn & Gravatt, 1995; Kulik, 1994; Niemiec & Walberg, 1987). These reviews provide (1) a clear description of predetermined criteria for selection of studies, showing that their data searches were unbiased and replicable; and (2) specification of how effect sizes are derived.

Niemiec and Walberg (1987) report an average effect size of .42 in their meta-analysis of 16 major reviews since the 1960s, 11 of which were meta-analytic reviews which assessed the relationship between CAI and achievement over a large number of studies (see Table 1). This is considered a moderate effect, showing that, on average, students who received computer-assisted instruction scored at the 66th percentile on tests of achievement compared to students in the control conditions without CAI who scored at the 50th percentile. However, they make a point that achievement may be differentially related to instructional level, as a number of studies reported student gains to be highest at elementary schools (.46) and lowest at college/university level (.26), with high schools in between (.32).

Table 1

<table>
<thead>
<tr>
<th>Instructional Level</th>
<th>Number of studies</th>
<th>Effect size</th>
<th>Percentile Gain over Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aiello &amp; Wolfle (1980)</td>
<td>Secondary, College</td>
<td>11</td>
<td>.42</td>
</tr>
<tr>
<td>Bangert-Drowns et al. (1985)</td>
<td>Secondary</td>
<td>42</td>
<td>.42</td>
</tr>
<tr>
<td>Hartley (1978)</td>
<td>Elementary, Secondary</td>
<td>33</td>
<td>.42</td>
</tr>
<tr>
<td>Kulik &amp; Kulik (1985)</td>
<td>College</td>
<td>101</td>
<td>.26</td>
</tr>
<tr>
<td>Kulik et al. (1985)</td>
<td>Elementary</td>
<td>28</td>
<td>.47</td>
</tr>
<tr>
<td>Samson et al (1985)</td>
<td>Secondary</td>
<td>42</td>
<td>.32</td>
</tr>
<tr>
<td>Schmidt et al (1985-86)</td>
<td>Special Education</td>
<td>26</td>
<td>.56</td>
</tr>
</tbody>
</table>

Note: Table excerpted from Niemiec and Walberg (1987)

These findings are consistent with Kulik’s (1981) model of instructional technology and student development. The model is based on the empirical evidence that achievement as a result of using CAI is inversely related to instructional level. CAI is most effective in the primary school. Student gains are somewhat less at the secondary level and lowest at the tertiary level. The use of CAI in special education seemed to be unusually effective with the largest effect size (.56). The authors speculated that for special education population, often mentally retarded or learning disabled, CAI may work more effectively for several reasons. It may be less threatening than traditional instruction; these students may benefit more than others from drill-and-practice and diagnostic procedures readily available in CAI.
In terms of cost-effectiveness, Niemiec and Walberg suggest that CAI is approximately twice as effective as peer tutoring. However, they also note that generalised costs and benefits are not well estimated and may fluctuate with time. Moreover, specific forms of CAI, student populations, and educational conditions have not been taken into account when estimating cost-effectiveness.

Table 2
Findings from Twelve Meta-Analyses on Computer-Based Instruction

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>Instructional Level</th>
<th>Number of Studies Analysed</th>
<th>Average Effect Size</th>
<th>Percentile Gain over Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen &amp; Dacanay (1991)</td>
<td>Health Professions Education</td>
<td>38</td>
<td>.46</td>
<td>18</td>
</tr>
<tr>
<td>Fletcher (1990)</td>
<td>Higher Education &amp; Adult Training</td>
<td>28</td>
<td>.50</td>
<td>19</td>
</tr>
<tr>
<td>Hartley (1978)</td>
<td>Elementary &amp; Secondary Math</td>
<td>33</td>
<td>.41</td>
<td>16</td>
</tr>
<tr>
<td>C. Kulik &amp; J. Kulik (1986)</td>
<td>College</td>
<td>119</td>
<td>.29</td>
<td>11</td>
</tr>
<tr>
<td>C. Kulik &amp; J. Kulik, &amp; Shwalb (1986)</td>
<td>Adult Education</td>
<td>30</td>
<td>.38</td>
<td>15</td>
</tr>
<tr>
<td>J. Kulik, C. Kulik, &amp; Bangert-Drown (1985)</td>
<td>Elementary</td>
<td>44</td>
<td>.40</td>
<td>16</td>
</tr>
<tr>
<td>Niemiec &amp; Walbert (1985)</td>
<td>Elementary</td>
<td>48</td>
<td>.37</td>
<td>14</td>
</tr>
<tr>
<td>Roblyer (1988)</td>
<td>Elementary to Adult Education</td>
<td>82</td>
<td>.31</td>
<td>12</td>
</tr>
<tr>
<td>Schmidt, Weinstein, Niemiec, &amp; Walberg (1985)</td>
<td>Special Education</td>
<td>18</td>
<td>.57</td>
<td>22</td>
</tr>
<tr>
<td>Willett, Yamashita, &amp; Anderson (1983)</td>
<td>Pre-College Science</td>
<td>11</td>
<td>.22</td>
<td>9</td>
</tr>
</tbody>
</table>

Note: Table excerpted from Kulik, James A. (1994)

Kulik (1994) conducted a meta-analysis of 12 CBI meta-analytic studies based on 546 individual studies. He reported average effect sizes ranging from .25 to .57, or between 10 and 22 percentile gain over a control group who performed at the 50th percentile. In addition, Kulik reports that students learned more in less time when they received computer-based instruction, and that students liked their classes more and developed more positive attitudes when their classes include computer-based instruction. These findings suggest that students who engaged in CBI performed significantly better than students who did not. However, CBI did not have positive effects in every area in which they were studied. As can be seen in
Table 2, achievement differed with educational level. At elementary level, effect sizes were greater, .37 to .40, or between 14 and 16 percentile gain, whereas at secondary and college level, effect sizes were smaller, .25 and .29, respectively.

The implementation of CBI in special education seemed to be most effective, with effect size .57, or 22 percentile gain. This may be due to the effect of individualisation of the learning process to accommodate students' learning needs. Similar to the findings of Niemiec and Walberg (1987), Kulik’s findings suggest that CAI is more effective in improving achievement of younger students and students with special learning needs.

In an updated meta-analysis on the learning effect of CAI based on 120 individual studies, Fletcher-Flinn and Gravatt (1995) examined the effect of a range of variables that related to treatments, methodologies, educational level, course content, and student characteristics. Their results and estimates were similar to previous reviews showing a learning benefit for CAI (see Table 3). The mean effect size of CAI for the year 1987-1992 was .24, suggesting that CAI students would outperform 60 percent of the students from traditional classes. However, no significant difference in effect size between educational level was found, though the effect for elementary level was slightly larger, .26 or 10 percentile gain when compared with .20 or 8 percentile gain at secondary and tertiary level. This pattern of results is inconsistent with Kulik’s model and those of other earlier reviews, the authors suggest that it may reflect a decline in novelty as computers become more commonplace in schools and at home. Effect size did not differ with CAI type. CAI implementation for drill-and-practice, tutorial, simulation, or word processing had a similarly moderate effect size of .23, .25 and .22 respectively. On the other hand, achievement was found differentially related to course content, with highest gains in mathematics, .32 or 13 percentile gain, and lowest in reading and writing, .12 or 5 percentile gain, an effect size considered not to be educationally significant.

Though it is expected that with more advanced hardware and software available there would be a significant increase in effect size over time, the findings of Fletcher-Flinn and Gravatt (1995) and Kulik and Kulik (1991) showed effect size fluctuated over years. Kulik and Kulik (1991) reported effect sizes of .24 for studies from 1966-1974, .36 for studies from 1974-1984, and .30 for studies from 1974 to 1985, and Fletcher-Flinn and Gravatt (1995) reported an effect size of .24 for studies from 1987 to 1992.
The implication of these findings is that effectiveness of CAI has not depended overly on advanced technology in courseware. Fletcher-Flinn and Gravatt make the point that studies where there was control for teacher and materials, and which were of longer
duration, and studies using pencil and paper equivalents of CAI, showed no learning advantage over traditional forms of instruction. They, therefore, suggested that what accounted for the typical learning advantage of CAI in this meta-analysis and others was the better quality instruction provided by CAI materials.

For the more recent meta-analyses, the scope of CAI studies tends to be narrower, though generally with similar moderate effect sizes. Examples of this are a meta-analysis of word processing in writing instruction (Bangert-Drowns, 1993), or a meta-analysis of CAI in elementary education (Ouyang, Gerlach, Bieger, & Mikkelsen, 1993).

Christmann, Badgett and Lucking (1997a) compared the academic achievement of students in grades six through twelve who received either traditional instruction or traditional instruction supplemented with CAI across eight curricular areas. Based on their 42 conclusions, an overall mean effect size of .21 was calculated, indicating that, on average, students receiving traditional instruction supplemented with CAI attained higher academic achievement than did 58.2 percent of those receiving only traditional instruction areas (see Table 4). The comparative effectiveness of CAI showed that the effect size for science subject was the highest, .64, and English the lowest, -.42, whereas maths was in between .18.

Table 4
Mean Effect Size in Subject Areas

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Number of Effect Size</th>
<th>Mean Effect Size</th>
<th>Percentile Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>4</td>
<td>-.42</td>
<td>-16</td>
</tr>
<tr>
<td>Math</td>
<td>15</td>
<td>.18</td>
<td>7</td>
</tr>
<tr>
<td>Music</td>
<td>1</td>
<td>.23</td>
<td>9</td>
</tr>
<tr>
<td>Reading</td>
<td>4</td>
<td>.26</td>
<td>10</td>
</tr>
<tr>
<td>Science</td>
<td>9</td>
<td>.64</td>
<td>24</td>
</tr>
<tr>
<td>Social Studies</td>
<td>2</td>
<td>.21</td>
<td>8</td>
</tr>
<tr>
<td>Special Education</td>
<td>3</td>
<td>.21</td>
<td>8</td>
</tr>
<tr>
<td>Vocational Education</td>
<td>4</td>
<td>-.08</td>
<td>-3</td>
</tr>
</tbody>
</table>

Note: Table excerpted from Christmann, Badgett, & Lucking (1997a)

Christmann, Badgett, and Lucking (1997b) also compared the academic achievement during a 12-year period (1984-1995) of secondary students across various academic areas who received instruction with or without using computers. They reported an effect size of .187, indicating that, on average, secondary students receiving computer-assisted instruction
over this period attained higher academic achievement than did 57% of those receiving only traditional instruction (see Table 5). This finding was similar to Fletch-Flinn and Gravatt’s (1995) reported effect size of .20 for studies at secondary school level from 1987 to 1992, showing a positive effect of CAI on achievement, but a smaller effect size. Furthermore, Christmann et al. (1997b) report a -.76 correlation between effect size and years indicating that the effect of CAI on academic achievement has declined during this period. This finding is contradicting the belief in the relationship between technological advancement and academic improvement (Mason, 1984).

Table 5

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Effect Size</th>
<th>Mean Effect Size</th>
<th>Percentile Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1</td>
<td>.88</td>
<td>31</td>
</tr>
<tr>
<td>1985</td>
<td>9</td>
<td>.61</td>
<td>23</td>
</tr>
<tr>
<td>1986</td>
<td>6</td>
<td>.39</td>
<td>15</td>
</tr>
<tr>
<td>1987</td>
<td>6</td>
<td>-.13</td>
<td>-5</td>
</tr>
<tr>
<td>1988</td>
<td>1</td>
<td>.23</td>
<td>9</td>
</tr>
<tr>
<td>1989</td>
<td>4</td>
<td>.20</td>
<td>8</td>
</tr>
<tr>
<td>1991</td>
<td>3</td>
<td>.31</td>
<td>12</td>
</tr>
<tr>
<td>1992</td>
<td>4</td>
<td>-.46</td>
<td>-18</td>
</tr>
<tr>
<td>1993</td>
<td>1</td>
<td>.16</td>
<td>6</td>
</tr>
<tr>
<td>1994</td>
<td>2</td>
<td>-.12</td>
<td>-5</td>
</tr>
<tr>
<td>1995</td>
<td>2</td>
<td>-.33</td>
<td>-13</td>
</tr>
</tbody>
</table>

Note: Table excerpted from Christmann, Badgett, & Lucking (1997b)

In summary, findings from both box-score reviews and meta-analyses show a general learning advantage for CAI over traditional instruction, though the level of effectiveness of CAI may vary with specific student population, course content and CAI type. However, it has yet to be established if this gain is simply an artifact of poor research design as Clark (1983) suggests. Another explanation is that the advantage stems from the generally superior quality of CAI materials, rather than from some intrinsic characteristics of the computer technology as a vehicle of instruction (Fletcher-Flinn & Gravatt, 1995).

In a recent review that did not employ meta-analysis, Sivin-Kachala (1998) assessed the effect of computer technology on learning and achievement by analyzing 219 individual research studies conducted from 1990 to 1997 across all learning domains and all learner ages. He reported that a) students in technology rich environments experienced positive
effects on achievement in all major subject areas; b) students in technology rich environments showed increased achievement in preschool through higher education for both regular and special needs children; and c) students’ attitude toward learning and their own self-concept improved consistently when computers were used for instruction. However, he acknowledged that the level of effectiveness of educational technology is influenced by the specific student population, the software design, the educator's role, and the level of student access to the technology.

2.3 Putting a face to meta-analyses: Specific reading and maths CAL software

There are a huge number of individual software programs available to assist reading and maths. The Los Angeles Times (1999) estimates that there are 700 programs for reading alone that are available to schools and homes. Thus, in this section it is proposed to include those that are in wide use (eg. those referenced by the U.S. Department of Education's National Diffusion Network, 1995) and have some reputable research associated with them. This means that the research has used valid measures of reading achievement, was conducted over a reasonable length of time and, generally, has been replicated.

The consideration of reading programs draws from the work of Schacter (1999) who reviewed selected reading programs from pre-kindergarten to 4th grade. Although he was less enthusiastic about the quality of research evaluating the reading technology programs he included, they are presented here as probably having better quality research associated with them than most. There is no implied judgement that they represent the best quality programs. Eleven reading technologies are discussed by Schacter and a further three programs presented that are currently in development, that is still being evaluated. However, as Successmaker, Josten’s and the Waterford Institute Program are discussed under a section of the report specifically devoted to ILS, they are not included here.

Some of the programs deal with a specific area of skill within reading. For example, Fast Forward (K-8) emphasises building oral language comprehension using computer altered speech sounds in an intensive training programme. The results of a large school-based study with 400 identified at-risk students K-3 showed significantly increased language comprehension and phonological processing ability of treatment over matched controls. Similarly, Daisy Quest, a pre-reading program for pre K-K designed to teach seven phonological awareness skills, led to significant gains for students (effect sizes .9 - 1.05).
A more broad ranging program, Wiggle Works, is an interactive CD-ROM (plus off-line material) that aims to teach phonetic awareness, letter identification and naming, story comprehension, and phonics skills. A study of 29 first grade classrooms (Schultz, 1995) showed that students using Wiggle Works made significantly greater gains on all reading (Iowa Test of Basic Skills sub tests of vocabulary, word analysis and language) and writing measures than comparison students. However, this is the only study of the program and, although individual differences in SES were controlled for, it is conceivable that school level effect of SES on reading growth rates could account for part of the gains found.

A major piece of software, IBM's Writing to Read has just been re-released in 2000. No data are available as yet. However, there is indication of the effectiveness of the earlier program in a review by Slavin (1991) of 29 separate studies, 13 of which were conducted by the Educational Testing Service. There was no evidence to suggest that Writing to Read had a positive effect on reading achievement of first graders but there was a moderate effect at kindergarten level (effect size .31).

Another piece of software with some of the features of an ILS like a reporting and management tool is Breakthrough to Literacy, designed for Pre K-2. It teaches vocabulary, phonological awareness, letter and sound recognition, word recognition and comprehension skills. From a study by Urrabazo (1998) of nearly 2000 students, Breakthrough students did not outperform those using other software or no technology. In fact, first graders using the program performed worse in all three schools studied. Until more research is conducted it is unfair to give the final thumbs down.

Accelerated Reader (AR) is a computerised reading management system that is designed for grades 3-11 and is part of a wider program, Reading Renaissance. Books are selected for their optimal reading level by the computer and students take multi-choice comprehension tests when they have finished a book. The computer scores and keeps/records to help teachers and students manage reading practice. Although several studies have examined the effectiveness of AR, only two have produced marginal evidence in favour of a positive effect on achievement.

Finally, the Academy of Reading by Autoskill was designed to help students increase their phonemic awareness, reading sub skills and comprehension. A basic part of the software is diagnostic tests that allow the development of an individualised program. There is little recent data on effectiveness. Older studies (Fiedorowicz & Trites, 1986, 1987, 1990) suggest that an Autoskill group significantly outperformed a control group on measures of
word recognition, phonetic knowledge of letters and syllables, paragraph reading, reading speed and reading comprehension. There was a reduction in reading errors and word recognition skills improved by 1.2 grade levels for the treatment compared to .3 for controls.

When one considers computer-assisted instruction in the area of maths, there are quite definite ways of regarding the available programs. Influenced by the work of Papert and colleagues at MIT, one area of use of the computer in maths is to create microworlds where the computer provides a simulation of a mathematical model that the user can manipulate and experiment with. There are a host of studies that look at whether the use of LOGO or a similar programming language helps students to construct their own understandings and concepts. It is not proposed to detail this research here. A meta-analysis by Liao and Bright (1991) of 65 studies looked at the effect of computer programming on cognitive abilities. They found a weighted effect size of .41, leading the authors to conclude that programming experiences provide “a mildly effective approach for teaching students cognitive skills in the classroom setting” (p. 262).

Table 6
Top Ten Mathematics CAI Studies

<table>
<thead>
<tr>
<th>Author</th>
<th>N</th>
<th>Grade</th>
<th>Duration</th>
<th>Topic</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cardelle-Elawar &amp; Wetzel (1995)</td>
<td>120</td>
<td>2, 4</td>
<td>15 weeks</td>
<td>Problem Solving Practice</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>2 Edwards (1991)</td>
<td>12</td>
<td>6-8</td>
<td>5 weeks</td>
<td>Transformations Microworld</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>3 Fletcher, Hawley, &amp; Piele (1990)</td>
<td>19</td>
<td>3, 5</td>
<td>4 months</td>
<td>Mathematics Drill &amp; Practice</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>4 Funkhouser &amp; Dennis (1992)</td>
<td>71</td>
<td>9-12</td>
<td>1 semester</td>
<td>Problem Solving Practice</td>
<td>No difference – ps + Treatment - math</td>
</tr>
<tr>
<td>5 Johnson-Gentile, Clements, &amp; Battista (1994)</td>
<td>223</td>
<td>5-6</td>
<td>8 lessons</td>
<td>Motion Microworld</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>6 Kiser (1990)</td>
<td>57</td>
<td>college</td>
<td>1 semester</td>
<td>Inequalities Demonstration</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>7 Thompson (1992)</td>
<td>20</td>
<td>4</td>
<td>9 days</td>
<td>Base-10 Blocks Microworld</td>
<td>No difference</td>
</tr>
<tr>
<td>8 Thompson &amp; Wang (1988)</td>
<td>20</td>
<td>6</td>
<td>2 weeks</td>
<td>Co-ordinate Graphing Microworld</td>
<td>+ Treatment</td>
</tr>
<tr>
<td>9 Varnhagen &amp; Zumbo (1990)</td>
<td>134</td>
<td>college</td>
<td>1 semester</td>
<td>Statistics Tutorial</td>
<td>No Difference</td>
</tr>
<tr>
<td>10 Zehavi (1988)</td>
<td>434</td>
<td>7-8</td>
<td>8 months</td>
<td>Graphing Microworld</td>
<td>+ Treatment</td>
</tr>
</tbody>
</table>

Note: Table excerpted from McCoy (1996, p.449)

There are other categories of use of computers in mathematics. In a review article of computer-based mathematics learning, McCoy (1996) concludes that the CAI research is
quite variable and includes microworlds, drill and practice and tutorial software. McCoy selects the top ten mathematics studies in programming, CAI and also mathematics tools (applications that allow graphing or symbol manipulation and construction or visualisation of figures). It is not clear how McCoy selects the top ten studies in each category except that she acknowledges that the studies reviewed are only a fraction of those that were considered, and that those chosen are "judged to be the most important to mathematics educators and researchers" (p. 439). Above is a table of the top ten CAI studies (Table 6), six of them are from the primary level. They include software that provides problem-solving practice, a transformations microworld, drill and practice, a motions microworld and a co-ordinate graphing microworld. All of the studies of these forms of CAI show positive effects for the treatment groups. Unfortunately, no effect size statistics are available to show either the strength of the effect or the relative effects.

2.4 Longitudinal and larger scale studies of CAL

There are a number of large scale projects and/or projects that have been operating for a considerable time. In this section, the overall findings from these projects with respect to learning outcomes are presented.

2.4.1 The Apple Classrooms of Tomorrow Project (ACOT)

The impact of interactive technologies in learning through the ACOT project was assessed by Baker, Gearhart and Herman (Baker, Gearhart, & Herman, 1994). There were five school sites across the nation and, over a period of five years, comparisons were made of student's basic skills performance to nationally reported norms; progress and achievement over time and teachers' teaching practices. Although ACOT had a positive effect on the latter two, there was no effect on standardised tests. ACOT students performed no better than comparison groups who did not have access to computers or to the teaching and learning reforms implemented in ACOT schools.

2.4.2 Computer Supported Intentional Learning Environments (CSILE)

The studies of Computer Supported Intentional Learning Environments (CSILE) where networked technologies are used to make working on a computer a collaborative and social enterprise, show positive results. Basically, in CSILE students conceive, respond to and reframe what is written and said over time on computer. They ask questions, search for answers and comment on other's work using computers. The study has run over more than eight years. Research in this time has shown that CSILE students do better than students in
control classrooms on various measures, including depth of understanding, reflection and also on standardised measures of reading, language and vocabulary (Scardamalia & Bereiter, 1996).

2.4.3 Wenglinsky’s national study of technology’s impact on mathematics achievement

This study (Wenglinsky, 1998) was concerned with mathematics achievement and the relationship between achievement on the National Assessment of Educational Progress (NAEP) test and key indicators of computer use. The four key indicators of computer use were (i) student access to computers in school for maths related tasks, measured by frequency of use, (ii) student access to and frequency of use of computers at home, (iii) preparedness of maths teachers in computer use, in terms of professional development in computer use, and (iv) the ways in which maths teachers and their students use computers, basically whether the use is for higher or lower order thinking skills. The study controlled for aspects of the social environment of the school including socio-economic status, class size, and teacher characteristics as well as using principal’s reports on lateness, absenteeism and morale. Thus, any relationships found represent the value added by technology. The report describes technology uses among 6,627 fourth graders and 7,146 eighth graders.

Using structural equation modelling, the study tested a model of how various technology characteristics might be related to various educational outcomes. The results show that technology does matter to educational achievement but the important moderator is the way it is used. Level of use does not matter but whether it is used for tasks employing higher order concepts is related to achievement. The analysis found that the students who used simulation and higher order thinking type software showed gains in maths scores of up to 15 weeks above grade level on the NAEP. An interesting finding was that students who used drill and practice technologies performed worse on the NAEP than students who did not use such.

This study, too, has methodological limitations, in that because there is no prior measure of maths achievement, the direction of effect is unclear. It may be that positive educational outcomes are conducive to certain types of technology use rather than the other way around.

2.4.4 The Computers Helping Instruction and Development

The Computers Helping Instruction and Learning Development (CHILD) study, which started in 1987, investigated the impact of computers on over 1,400 students and their
teachers from nine Florida elementary schools (Kromhout & Butzin, 1993). Results of the study showed positive and statistically significant changes in standardised test scores for all participating grades and schools. The largest effects appeared for students involved in the program for more than one year.

2.5 The special case of Integrated Learning Systems (ILS)

The integrated learning systems in use are the product of three to four decades of development. In many different countries they are one of the major ways that computers are used in schools today. From figures in the 1992 IEA study, it is estimated that 20 per cent of all U.S. elementary schools have ILS installed (Becker & Hativa, 1994). ILS constitute a major spending category for educational technology. In the United States, in terms of software expenditure, one half of the current dollar investments are going to ILS companies (Bailey, 1993). These systems are also one of the most successfully marketed (Becker & Hativa, 1994). It should be noted that "a great many ILS-implementing schools are places where the pressure to raise test scores is strong, either because of high expectations from parents or because the students are demonstrably failing to master the basic arithmetic and language arts skills during their elementary years (Becker, 1994).

The theoretical underpinning of ILS products is in neo-behaviourist learning theory, whereby tasks are automatically selected for an individual, practice given and feedback offered in order to modify behaviour. It is a mastery approach whereby tasks are ordered according to an expert's opinion of hierarchical difficulty (Rodrigues, 1997). Learning is perceived as an individualised activity between the student and the material. An ILS offers differentiation of curriculum content for a learner but not differentiation of methodology. Although there is an increasing body of research on ILS, the difference between products, the different learning contexts, different modes of integration and different research designs make the interpretation of findings a complex task.

There exists a review of ILS in the US (Becker, 1992) and, more recently, evaluations from a large study in New York (Miller, 1997a) and in West Virginia (Mann, Shakeshaft, Becker, & Kottkamp, 1999). Then there are findings concerning ILS use from the large UK three phase study (BECTA, 1998; NCET, 1994, 1996). Reasonably designed smaller scale studies will also be reported. The emphasis in this section will be on the findings relating to learning outcomes in literacy and numeracy. Particular attention will be paid to references to SuccessMaker in these studies (e.g. in the U.K. evaluations SuccessMaker was the major form of ILS). The impact of ILS on attitudes and behaviour is not emphasised or considered.
a proximal indicator because, according to Wood, Underwood and Avis (1999), there is little evidence from the U.K. studies that learning outcomes were influenced by attitudinal factors.

The different bodies of research will be considered and commented on, in turn. As the current New Zealand use is largely concerned with low achieving populations, any research specifically addressing this factor will be drawn on and presented in a separate section. It is not appropriate to proceed in the traditional literature review way of integrating findings across studies using themes as differences in the systems, the sites where they are implemented and research designs preclude this. The summary will attempt to identify common findings and themes.

2.5.1 An early review of ILS in the U.S.

A major review of thirty evaluation reports of computer-based integrated learning systems in elementary and middle grade schools was conducted by Becker (1992). There were several different ILS included in this review. The review is important because it highlights difficulties inherent in drawing inferences from data. For this reason, the review is worth examining in detail. The purpose of the review was not to compare systems as Becker cautions that the sites and research designs employed to study the effect of one company's product are not necessarily comparable with those used to study another system. Leaving aside any debate over the desirability of a single vendor, centrally managed, diagnostically prescriptive, networked system for addressing curricular goals, Becker asks, "Do these systems work? Do students make more progress using them than they would have, had their time been spent learning the material in more traditional ways? Is this progress on important curricula goals?" Becker points out that the studies he used represent "a diverse array of district-disseminated and vendor-disseminated evaluations and independent research projects <but> they cannot be really seen as an unbiased sample of current implementations of integrated systems" (p.4).

There are a number of issues that Becker identifies as important to consider in interpreting the conclusions from the studies. These include selective dissemination as a barrier to fair interpretation of existing data on ILS, as well as poor evaluation designs, compounded by inadequate data collection; poor data analysis, and insufficient description of the implementation of the program, including the context, and conditions and environment where the programme was used. Particularly important in this latter case, is whether the students using the program spent more time on the subject than did those in comparison groups. Regarding design and data collection, there is the difficulty of finding an appropriate
comparison group, particularly where there are problems arising from inadequately normed standardised tests. There is also the issue of whether this standardised test adequately reflects instructional experiences.

To make sense out of the diversity of statistics contained in the reports, Becker used effect size, a measure that converts the different types of information about effects into a common measure of effectiveness. This is the difference in performance of students using ILS and a comparison group, expressed as the proportion of one standard deviation of individual scores in the performance group. There still remains the issue of what is an effect size that would be considered a reasonable gain? Given that, for various reasons, there is a national upward drift in test scores in the US, Becker (1992) argues that to be considered a reasonable gain for a one-year effect size is .3; over three years, an effect size of .47 is necessary. He says that given an appropriate comparison group, these would represent a successful large-scale instructional intervention. It should be noted, as discussed above, that other researchers have a different view of what might constitute a low, moderate and high level of improvement in terms of effect size.

The analysis reported covers 16 evaluations of software from the World Institute for Computer-assisted Teaching (WICAT); 11 evaluations of the Computer Curriculum Corporation software (CCC and SuccessMaker are used interchangeably to refer to the software program produced by this company. It appears that SuccessMaker is a more recent term) and four of Jostens (ESC) and a multi-vendor evaluation done in New York City. The data Becker presents includes the location and the system used; who conducted the evaluation and produced the report; brief outline of type of research design and comparison group used; how many users supplying data; grade level; socio-demographic characteristics of the school and extent of experience on ILS. Also included is data from the evaluations regarding how the evaluation described the effectiveness (how achievement measured; what comparisons among students were made and what was found). Becker translates the data provided by the original evaluator’s comparison into an effect size statistic. He notes what assumptions were made and provides high and low estimates in such cases. His table summarising these aspects for the CCC ILS system is presented in the table in Appendix 2.

Becker’s conclusion is that there are widely varying effect sizes arising possibly from different conditions of the studies; different software packages and the different methodologies used to design and conduct the studies. This aside, there is typically a modest to small positive effect (e.g. median effect size of WICAT studies through vendor .17; CCC through vendor .40; CCC independent .15; Jostens through vendor .0 and independent
Students generally do somewhat better using ILS than they would be expected to do and sometimes much better although it is not clear the conditions under which these advantages accrue. The more detailed findings regarding CCC are presented in the following paragraphs.

With respect to CCC, his conclusion was that some studies substantially over report effectiveness. The largest reported effects came from four evaluations supplied by CCC (Aiken County, Milwaukee Public Schools; Corpus Christi Parochial Schools and Omaha Westside Community Schools). Each was a year long study with an hour and a half a week in maths and reading. Effect sizes ranged from .6 to 1.6. There were two analytical problems noted for these studies. The first was that the vendor evaluators eliminated cases from data analysis that showed large declines, arguing that singular conditions made such post-test scores atypical of an individual's performance. However, students who showed unusually large gains from their pre-test scores were not similarly excluded. Such asymmetry was defended on the grounds that the treatment could have caused the large gain but not the drop. This clearly ignores the other explanation of unusual circumstances on the test day for small numbers of individuals. However, correcting for this deletion of negative outliers only reduces the effect size substantially in one study. The other problem was that, in two of the four studies supplied by CCC, there was a very high attrition rate, that is data are only supplied for about 60% of the students enrolled on the program. There is no attempt to show if these drop outs differed systematically from those who stayed. It should be noted that these large effects are not replicated in other studies of CCC.

Becker, similarly, deflates the findings from the Calvert County, Maryland study of CCC (Austin, 1988 cited in Becker, 1992), pointing out that the district most commonly compared the most recent scores with those obtained three years prior to the use of ILS. However, both district and state mean grade equivalent scores were rising in the period 1980-1986, prior to ILS. Becker shows that at all three grade levels reported in the ILS study, district scores are worse as of 1988 than they were at the onset of the ILS program, after state-wide gains in mean grade equivalent scores are taken into account. Another way Becker presents the outcome data is to compare whether the grade cohorts that had several years of CCC made greater test score gains than the earlier cohorts that did not use the program, again using gains state-wide of the same cohort as a comparison point. Such an analysis showed that between third and fifth grade that the four cohorts who had CCC did slightly better than two earlier cohorts (estimated effect size .12). However, between fifth and eighth grade two cohorts with two to three years experience of CCC showed less improvement than cohorts who had less than half a year on CCC (estimated effect size -.25).
Even the slight gain shown by the younger students may be explained by demographic change as younger, more educated suburban families moved into a largely rural, older district. The conclusion is that a re-examination of the data from Calvert County, of both the district level cohort test scores and the longitudinal data on individual students, suggests a conclusion opposite to what was at the time promulgated. The conclusion is that test scores rose no faster than they would have in the absence of an ILS.

Because all of the above studies of CCC involved the use of an artificial comparison group, which has numerous problems, Becker also reports two studies involving CCC where the allocation to ILS-using and traditional instruction was randomised. One of these studies was in Los Angeles (Ragosta, Holland, & Jamison, 1982, cited in Becker, 1992) conducted by the Educational Testing Service, reported large effect sizes on standardised achievement tests for mathematics computation, accumulating to .7 over three years. However, effect size for maths concepts and applications was close to zero, making the overall maths score lower but still substantially positive. There were also modest effect sizes for reading and language. In addition, there were effect sizes reported for curriculum-specific tests. These Becker describes as huge, then questions whether, as they were drawn from computer-based activities, the test was balanced as the material may not have been experienced by the non-ILS group. Perhaps more significant is the fact that, according to Becker, the ETS used a non standard and upward-biasing standard deviation measure in all effect size calculations (ie. an effect size based on a standard deviation of residual post-test scores after controlling for pre-test scores). Because pre and post-test scores are correlated, the distribution of residual scores is more compact than the actual post test scores. When the effect sizes were recalculated using a more conventional and standard definition of standard deviation, the maths effect sizes reduced by 25% and the language by 40% (ie. .26 for maths computation over one year and .59 over three; .17 and .27 for total maths and .12 and .14 for language arts over the year and three years, respectively). The other study, The LaFayette Parish experiment, found an effect size of .2 for mathematics for one year, about the same as the adjusted Los Angeles result.

Perhaps the most important point to be taken from Becker's careful work is that methodological weaknesses render a number of studies invalid and unusable. Other studies that he worked with have data that can be reanalysed to correct or allow for weaknesses in design. When these factors are employed, the median effect size for CCC is .40 for studies supplied through the vendor and .15 for studies conducted by the local education authority or an independent researcher and .17 for the randomised design studies. This points to the need to be wary of analyses that incorporate studies of widely varying quality without any
reanalyses. It also points to the need to not accept uncritically the studies supplied by vendors.

2.5.2 Recent large scale US evaluations

New York. The New York City Public Schools Integrated Learning Systems Project which began in 1989 was reported in an issue of the International Journal of Educational Research in 1997 (Miller, 1997a). The evaluation was conducted by a team based at Brigham-Young University. The report is accompanied by an independent meta-evaluation, a critique of the evaluation in terms of a comprehensive set of professional standards. Both were funded by an anonymous party independent of the evaluation team and by the project's sponsors. However, the authors of the meta-evaluation note "The evaluation of the WILS was not a true field test since there was heavy involvement by the Waterford Institute throughout the study" (Finn, Stevens, Stufflebeam, & Walberg, 1997, p.169). By this they mean that Institute personnel provided hands-on support features that would be difficult to replicate. The Waterford Integrated Learning System (WILS) was the main component of the program.

Ten schools participated in the project, chosen because of "fiscal and attitudinal readiness" (Miller, 1997b, p.94). The profile of schools is that they consisted almost entirely of minority students from lower socio-economic status homes. About a sixth had limited English language proficiency. The majority of students scored below the 50th percentile in standardised tests of reading and maths. Only a quarter of the students scored above the 50th percentile in reading.

As those who conducted the meta-evaluation noted "The report is voluminous compared to its message- essentially no sustained, significant effects of the Waterford Integrated Learning System (WILS) on student test scores and attitudes" (Finn et al., 1997, p. 159). The lack of effects was pervasive, applying to low achievers as well as others. In each school, pre ILS NCE scores were compared with those following the innovation. If the WILS produced any effect in terms of an increase in test scores, then the percentage of positive differences should increase for the year in which the installation occurred and those following. The results of an analysis of changes in maths and reading scores across adjacent years for grades 3 through 5 in treatment schools give no basis for asserting a positive contribution of ILS. The mean differences between the scores for treatment and comparison schools were also computed. Although the finding was in the right direction it was not significant and, indeed, the difference between comparison schools was comparable.
to that between treatment and comparison schools. There was also an analysis using planned comparisons of the scores of students who stayed at the same school for two consecutive years, to get around the problem of student turnover. Again the total number of significant comparisons for these students was no different to that for students in control schools. No matter what form of analysis was undertaken, no significant effects were found. Finally, summarising the mean effect size for grades 3 and 5 at all schools in the pre-treatment and treatment years, shows that there was evidence of effectiveness for maths though the mean effect sizes in the treatment years averaged .2.

The meta-evaluation found that there was a reasonable, though not close, match between WILS schools and comparison schools and that the evaluation was sound, making good use of multiple sources of information. It was noted that it was appropriate for the report to conclude that the factors contributing to the failure of the WILS were unknown, complicated and, possibly, numerous. The meta-evaluators note the fact that most participating schools displayed institutional rigidity, resistances, mishaps and incompetence and that there was high student and teacher mobility. An intervention is always subject to aspects of the institutions and their functioning and environment. In four of the ten test schools, the intervention may not have been in place long enough to establish efficacy. They also note in reading, in particular, the discrepancy between the curricular and pedagogical strategies that New York City teachers were being asked to employ in classes and the approaches built into the WILS. An important conclusion is that the approaches of the innovation should most likely mirror those that the school system is otherwise employing. For an intervention to realise its potential, it probably needs to be closely integrated with "other aspects of instructional process and institutional function" (Finn et al., 1997, p. 172).

The authors of the meta-evaluation feel that the report was silent on the key question of whether the New York City Public School Board of Education should adopt WILS widely. "Instead of saying that the Board might be well advised not to expand its investment in WILS, late in the report, there seemed to be an undertone that WILS was sound and the reason for its failure to demonstrate effectiveness was inadequate implementation. We think there was a need to consider directly and validly whether further installation of WILS in public elementary schools would be warranted, given its considerable costs and lack of demonstrated effects" (Finn et al., 1997, p. 169).

West Virginia. In West Virginia, the ILS it was known as the Basic Skills/Computer Education Program (BS/CE). The ILS focussed on teaching the State's basic skill goals in spelling, vocabulary, reading and mathematics. The software was chosen by the schools
from IBM or Jostens Learning. These both met the State's instructional goals but differed in their approach. Professional development was an integral part of the program. An analysis, commissioned by the Milken Exchange on Education Technology, was undertaken by Mann, Shakeshaft, Becker and Kottkamp (1999) of a representative sample of 950 fifth-grade students' achievement (in 1996-7) from 18 elementary schools across the state, schools who had been participating in the program since 1991-2. The bottom line of their findings was that the more students participated in the BS/CE, the more their test scores rose on a national test, the Stanford 9 achievement test. Although all students' test scores rose, the neediest students (classified as those who do not have computers at home) rose the most. The BS/CE technology regression model accounts for 11% of the total variance in the basic skills achievement scores of the students. The three components of the model were access to technology, attitude, and training. The writers of the report argue that this is 11% of the 30% that they believe is available to the schools to influence. This line of argument stems from Coleman's (1966) notion that over 70% of the variation in achievement for each group is variation within the same student body which is interpreted as family and other home and background factors. Thus, 11% as a portion of 30%, they see as considerable!

In the Preface to the report, the Director of the Foundation cautions that the system was designed over a decade ago and was limited to the then available technology. Like other ILS, it is based on a pedagogy that "makes little use of project-based learning and other constructivist curricula approaches that are the leading edge of technology today" (Mann et al., 1999, p.3).

2.5.3 The United Kingdom ILS evaluations

This was a three phase project with several independent (of the software vendors and of each other) research teams (BECTA, 1998; NCET, 1994, 1996). The phases each had control groups and independent tests to assess outcomes. There were different criteria for assessment of outcomes across the three phases. Where there is convergence in results this helps to assure of their validity. It also helps where there are differences to look at aspects of the scope and generality of effects of ILS on learning. In the U.K. in the initial phases, only two ILS systems were evaluated, SuccessMaker and SIR's Global Learning Systems. In phase three Jostens was also included, but by far the greatest number of students (62%) were in schools using SuccessMaker.

There were 23 schools involved in Phases one and two; studies where the usage of the ILS was six and three months, respectively (NCET, 1994, 1996). These schools were not
randomly selected but self-selected and probably more ILS prepared than the norm. The matched controls were within school but, in controlling for school effects, it is possible that non-ILS students were affected by the intervention.

The most general finding from this early phase was of significant and lasting effects of SuccessMaker on achievements in basic maths skills (NCET, 1994). In phase one ILS students from 12 schools outperformed controls in maths (effect size .45). In phase two, with a sample of new learners, over a shorter time (2-3 months), the effect was smaller (.1). The results for SuccessMaker in English were not clear in Phase two (NCET, 1996). The assessment outcome measures were focused on basic or core knowledge and skills in English and maths. There was no assessment in phases one and two of more general conceptual understanding, knowledge application and/or performance in examinations. The main issues therefore concern generalisability of evidence for gains to other methods of assessment and across schools and students.

What was clear was that ILS are not a coherent class of systems and the same system could have differential effects across subject areas. Another important finding from these early phases was that, although some schools were making use of the normative data provided by the ILS management system, there were reservations about its validity. This is because the database from which the system makes assessments about initial knowledge, placement or progress evaluation was not known and this particularly applied to SuccessMaker. Thus, any measures derived from such systems are not so much indicators of developmental progress as of what was taught, when and how.

In Phase three, the major aim was to see how the results from ILS generalised to different performance measures, what the Final Report calls “unpacking what is meant by the phrase ‘effective in producing worthwhile gains in literacy and numeracy’” (BECTA, 1998, p. 16). In this phase, learning was evaluated against standardised tests and in terms of the impact on external examinations. These included curriculum-oriented tests of maths and English; a test of cognitive ability and the Key Stage 3 examination and GSE examination grades (both national, external examinations). Years 5 and 8 received the curriculum-oriented test from one group of researchers, while another group of researchers considered the examination results of years 9 and 11.

For the curriculum oriented testing study, a larger sample than in previous phases was involved for a comparison of ILS and non-ILS students from 11 and 19 different schools, respectively, in the primary sector and 25 ILS and 23 control schools in the secondary
sector. In Year 5, 216 ILS students were recruited and between 186 and 197 completed the performance measures with 675 control students of whom between 61 and 284 completed both rounds of testing. In Year 8 between 1345 and 1528 ILS students completed testing along with between 191 and 276 controls. There were some Year 6 students without controls (between 296 and 328 students completed both pre and post tests). Given the initial sample size and the attrition rates, it means that some sub groups are small, remembering the students are distributed among three systems. In addition, there was a small-scale longitudinal study of ILS and control students involved in earlier phases to assess the extent to which any differences found between ILS and control translated to examination performance.

In attempting to find out what learning outcomes ILS supports, information was used to construct a model to find the best fit between factors such as student background (age, gender, ethnicity, socio-economic status) and variation in an overall measure of student performance on the tests used. Whether or not a student used ILS was included to see if this increased the model's predictive power in terms of learning gains. As this phase of the study was not a matched design, samples of year 5 and 8 were used to establish statistical models that were then used to test the effects of ILS. The effect tested was whether ILS students, in general, fared better than non-ILS controls and to see whether gender, ethnicity, and SES interacted with this.

The results for Year 5 show a small but statistically significant difference between the gain scores for ILS and non-ILS schools. However, the gain scores for SuccessMaker students were about the same as those from schools without ILS. Year 6 students showed large gains in maths, perhaps explainable in terms of the fact that Year 6 had just been prepared for their key stage assessment test. However, the gains in English were not comparable and this tends, therefore, to weaken this explanation. The result tended to support the Phase two finding that ILS (specifically SuccessMaker) supports learning in maths more effectively than in English for this age group. The Year 8 group, similarly, had a small but significant difference in favour of ILS. The effects for maths which were moderate in Phase one (.45) were negligible in Phase three (less than .1). There was evidence, however, of significant difference across schools although the data obtained do not provide anything conclusive in terms of a factor or factors that might explain this.

As is often the case in evaluation studies, the researchers have to work with what is available rather than what is ideal. For example, in Phase three, the evaluators wanted to work with mature users of technology to overcome the difficulties of trying to assess learning
effects while there were still shake down issues in terms of integrating the system. However, to obtain a reasonable sample size, they included new users. The ideal design of random allocation of students/schools to the ILS condition was not possible.

The other performance measure to be considered was the exam performance of 416 ILS and 599 non-ILS Year 9 students from eight schools. The statistical model used to assess added value was based on 1372 students and was part of data collected previously. After attrition, there were 357 ILS students who had Key Stage 3 data and 283 for whom there was MidYIS data. In Year 11 there were six schools involved with a total of 214 ILS and 415 controls and GSCE grades were the performance measure. The statistical model used to predict expected outcomes was based on over 37,000 students who had participated in a previous national project. So, although this research team used the same statistical modelling techniques as for the younger (Years 5 and 8) population, the model was derived from an extensive database. Thus, the patterns of relationships found in that data tell us what we could expect a student, with a given profile, or history, to achieve in state examinations. Such an approach is able to support inferences about the likelihood that a given individual student's scores are within, above or below the normally expected range of scores for his/her profile group. In addition, because there was within school control, differences attributable to school effects can be accounted for.

The general pattern of results of the effects of ILS on student's performance over all 14 schools involved at the two year groups was consistent. As the Report of the team records, "Under the challenging conditions <of exam performance> it was difficult to find anything but an apparently negative effect on achievement. The ILS samples achieved lower than the model predicted grades. Where non-ILS pupils from a school had positive residual gain scores, their ILS peers had less positive ones. Where controls had, on average, negative gain scores, ILS had more negative ones" (p.18). In terms of effect size, for maths they ranged from -.04 to -.62 for Year 9 and -.03 to -1.29 for Year 11. Although the negative effects are relatively small, they are consistent and significant. However, the use of the word 'apparently' in the research team’s conclusion serves as a reminder that as assignment to ILS and control groups was not random, there may have been some 'biasing' factors. In this case, whatever the non-ILS students were learning prior to their exams, was more beneficial than ILS experience in terms of these outcome measures. In addition, this research team looked at the relationship between assessments of progress provided for work on the SuccessMaker system and examination performance. They found a modest correlation of a similar magnitude as that between exam performance and socio-economic factors. The
researchers caution that the predictive power of ILS normative measures in terms of external measures of achievement is not strong.

When a comparison is made of the findings from Phase three with those of earlier phases, the major discrepancy concerns the impact of SuccessMaker on maths achievement. The table below (Table 7) summarises the sample profiles and effect sizes for maths in the three phases. These range from highly positive, through neutral to negative. There are a number of factors that individually or in some combination may explain the conflicting findings. These include differences associated with selection of schools and samples; differences in test administration; differences in statistical procedures used by the various teams to make comparisons and differences in the outcome measures used to assess learning. It appears that the latter is the most plausible explanation, namely the nature of the performance assessments and how these relate to the learning goals best supported by ILS. It seems likely that the tests used in Phases one and two placed emphasis on basic maths skills of the procedural kind whereas the tests in Phase three were of the sort that demanded more reasoning and translation. "The implication is that ILS, particularly SuccessMaker, are good at supporting the learning of basic or core knowledge, skills and procedures but less effective in promoting the knowledge and practices that go into the interpretation or translation of tasks into recognisable problems for solution" (BECTA, 1998, p. 20).

When considering the U.K. evaluations as a whole, it needs to be borne in mind that across the three phases of the research there were many variations in design and the way the evaluation was carried out. There were apparent differences in findings across the three phases. How far these differences in findings stem from methodological differences is not easy to determine. In addition, the contexts for implementation may have been different with the extent of teachers' experience with ILS and with integrating technology varying widely. There was evidence for school-based effects in relation to measures of learning outcomes. But, there is no consistency in implementation to allow interpretation of these effects. Also, schools continued to change as they experimented with integration so, again, identifying salient features and establishing models of implementation was not possible. Where there were no control groups internal to the school, it is not possible to rule out that school-based differences are reflections of general differences in the achievement of students from different schools. They may or may not reflect differences in the use of ILS. On the other hand, when there are no external controls, it is unclear whether any effects within a school come from positive effect of ILS on students or involve some unexplored negative impact of not being on ILS on their controls. Finally, without random assignment of schools and
students, it is impossible to be sure that effects found are not due to some factor or factors undetected in the data analysis but which cause systematic bias in the choice of students for ILS.

Table 7
Summary of Sample Profile and Effect Sizes for Mathematics: Phases 1 - 3

(a) Phase 1 and 2 - Leicester

<table>
<thead>
<tr>
<th></th>
<th>Primary</th>
<th></th>
<th>Secondary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>1</td>
<td>39</td>
<td>5</td>
<td>180</td>
</tr>
<tr>
<td>Phase 2</td>
<td>6</td>
<td>505</td>
<td>8</td>
<td>554</td>
</tr>
</tbody>
</table>

(b) Phase 3 - NFER

<table>
<thead>
<tr>
<th></th>
<th>Year 5</th>
<th>Year 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILS sample</td>
<td>11 schools, 186 - 197 pupils</td>
<td>5 schools, 1345 - 1528 pupils</td>
</tr>
<tr>
<td>Controls</td>
<td>19 schools, 61 - 166 pupils</td>
<td>27 schools, 191 - 244 pupils</td>
</tr>
<tr>
<td>CCC &amp; global maths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect size</td>
<td>-.10</td>
<td>.10</td>
</tr>
</tbody>
</table>

(c) Phase 3 - DURHAM

<table>
<thead>
<tr>
<th></th>
<th>Year 9</th>
<th>Year 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILS sample</td>
<td>8 schools, 416 pupils</td>
<td>6 schools, 214 pupils</td>
</tr>
<tr>
<td>Internal controls</td>
<td>8 schools, 99 pupils</td>
<td>6 schools, 415 pupils</td>
</tr>
<tr>
<td>External controls</td>
<td>1300 pupils</td>
<td>37,000+ pupils</td>
</tr>
<tr>
<td>Effect size</td>
<td>-.40 to -.62</td>
<td>.00 to -1.29</td>
</tr>
</tbody>
</table>


Data for CCC and Global schools contributing ILS and control pupils in mathematics.


The effect sizes refer to results from all ILS groups.


Hulme et al. (1998).

Obviously, students learn on ILS. There is some evidence that time on the ILS system contributes to the measures of learning outcomes used, particularly with respect to acquisition of core knowledge and skills. The issue is rather what and how they learn. The question posed in the conclusions to the final report was, "Is what an integrated learning system teaches, worth knowing?" As is pointed out, one way to do this is to compare the content against the demands of the National Curriculum. They conclude this is neither easy nor sufficient but instead say we need to ask searching questions about how learning tasks are represented, diagnosed and taught. "If one were to take performance on Key Stage 3
and GCSE results as an index of this (i.e. as reliable, valid and summative measures of what children have learned about the National Curriculum), then the evaluators have failed to provide any evidence that ILS is 'effective' or 'efficient' in supporting the development of either numeracy or literacy on this index" (BECTA, 1998, p. 31).

There is a further issue to consider arising from the British study and that concerns the fact that if ILS are shown to support the learning of basic skills, then why are there no effects showing through at GCSE level? There are a number of potential answers to this finding and each raises issues about the use of ILS. One possible explanation is that, at the level of examinations, the knowledge and skills that ILS support is taken for granted so such knowledge does not contribute to variation in exam scores. The implication of this is that the task of exploiting students' basic skills and helping them to apply them in novel situations and tasks will be achieved in contexts other than where ILS is used.

A second possible explanation is that, as one U.K. evaluation team suggested, there is a misalignment between the content of ILS teaching and what is examined. In other words, ILS may lack curriculum validity and this becomes more marked at later stages than early on when core skills are being acquired. If one accepts this explanation, then the use of ILS will only be helpful in a limited range of curricular topics. A third explanation could be that, although the basic skills taught by ILS are age and curriculum appropriate, additional teaching and learning is necessary to make them relevant and applicable to situations where knowledge has to be applied and used flexibly. The implication is that teachers need to bridge the gap between what a system teaches and the application of that knowledge. Yet another explanation is that the representation of knowledge and of what it is to know is too limited and may tend to concentrate on single ways of representing and solving problems. For teachers, the challenge is to diagnose the limits of a system and fill the gaps, particularly in relation to the task of making connections and explaining relationships between different ways of handling, representing and working on situations.

Finally, the authors of the Report suggest that it may be that ILS propound a model of learning to the learner which inhibits the development of attitudes, concepts and processes needed to become adaptive, creative and flexible learners. Thus, the process of learning procedures to solve well-defined problems, hinders the process of learning to conceptualise situations for oneself. This would include things like how to set and evaluate goals, how to search for and evaluate information etc. in order to make decisions and problem-solve. If this is accepted as an explanation, it may be that other forms of technology may contribute to this learning rather than ILS.
2.5.4. Other studies of ILS

A recent study of an integrated learning system, Learning Expedition (Taylor, 1999) suggests that there are some gains possible in the mathematics area. Learning Expedition is somewhat different from many ILS in that it allows a high level of teacher control over the range of topics for an individual student. In this study of 11-13 year olds in one school, time spent on the ILS added significantly to the predominant explanatory variable of prior level of achievement (on the National Foundation for Educational Research verbal, non-verbal and quantitative tests) in explaining results on the end of year maths exam.

The only published New Zealand study was of the use of SuccessMaker in a secondary school (Parr, 1997). Although students made sizeable, rapid gains, particularly in maths, on the system administered measures, this did not generalise to standardised achievement test (PAT) or school based assessment. The greatest gains were made by those who spent the most time on the system on a regular basis. These were the lowest achieving students. It should be noted that these students had a very favourable student-teacher ratio while they worked on the ILS.

2.6 Ability and ILS findings

An instructional program cannot be assumed to be equally effective for all groups of students. Individualised instruction is touted as a major feature of ILS. It may be that higher ability students can proceed at their own pace, ahead of where whole class instruction may have taken them, but it is equally clear that lower achieving students often cannot make progress using computer software tutorials and drills without the help of other people. The effectiveness of a seemingly individualised and automated system as in an ILS, is dependent on what Becker calls the mindfulness with which it is used (Becker, 1994).

The British studies have some remarks regarding whether ILS may operate differentially with respect to students of differing ability. The reports noted that the results from Phases one and two indicated that schools were likely to view ILS as particularly suitable for low performing children. However, the results showed that ILS gave neither special benefits nor disadvantages for such students in terms of learning outcomes. In Phase three, a research team considered special needs children using AutoSkill. Again, there is no strong evidence of any differential impact on these students.

Of interest is the finding by Becker (1994) of a curvilinear pattern of effects in terms of the relationship between outcomes and ability. In his study of two computer-based integrated
learning systems (one from Computer Networking Specialists and one from Josten's Learning Corporation) at two elementary schools, Becker employed a research design that used same classroom, pre-test matched controls. This was so that effects of three of the major contributions to student achievement, namely, prior test scores, teacher effects and time effects were controlled. Results showed that the students at the top and bottom of their school's prior achievement distribution did better when using ILS, and students in the middle did better with only traditional instruction given to them. Neither school produced positive results for their total enrolment across reading and maths, only certain categories of student profited. Mostly it was the higher achieving students who benefited, then next it was those in the bottom third. In one school, where the Josten's program was used, the middle group of students had the lowest effect sizes for all three reading and language tests. However, in maths, students in the middle had moderately positive effect sizes on all maths measures, while the low group generally had negative effect sizes. The results for high students were mixed, substantially positive on a more conceptually difficult test and moderately positive on computation but negative on norm referenced test of concepts and applications. At the other school, there were few sub group differences in reading but the maths results were similar to the reading results at the other school, whereby both high and low groups had very positive effect sizes while the middle group did better in the control treatment, normal teaching.

These findings are corroborated by Osin et al (cited in Becker, 1994) who found that high and low achievers progressed faster than middle level students. The explanation is that classroom instruction is more often pitched to the middle and individualised work is better able to address the needs of extreme aptitude groups.

An early review of computer-based instruction (Swan, Guerrero, Mitrani & Schoener, 1990) aimed to investigate its efficacy for providing basic skills remediation in maths and reading for educationally disadvantaged student populations. (Note: Although the term computer-based instruction is employed in the article, most of the programs reviewed are ILS, including CCC, WICAT, ESC, PLATO). In this review 13 programs based in 26 elementary and secondary schools in New York City were considered. The design was a within subjects one, where students in need of remediation were selected by the schools and their classes scheduled in compliance with the particular systems requirements. The test scores from the Degrees of Reading Power (DRP) and Metropolitan Achievement Test (MAT) were compared for the years 1987 and 1988 (overall effect size calculated by dividing the mean differences by their standard deviations). Note that there was no correction for any upward drift state-wide in these tests in the period.
The overall conclusion was that such programs were an effective way of delivering remediation, equally effective in reading (overall effect size .8) and maths (overall effect size .9). The effect sizes decreased with grade level going from 1.1 for reading in elementary school to .3 for junior high and in maths from 1.2 to .4). Thus, there was an inverse relationship between instructional level and achievement gains. Further analyses showed that there were differences between different groups of students, namely students at different levels of remediation from special education students through to what are termed general education students. There were significant differences between groups and this was particularly marked at the lower grade levels. Special education students benefited most at all grade levels and remedial and general education students least at most grade levels. However, this is confounded by the fact that different ILS were used at different grade levels and these may have been differentially effective both in general or at different grade levels.

2.7 Summary of ILS

What is obvious about the findings from the studies of ILS is the lack of consistency. Both across studies and even within one study, there is variability. For example, in the four main investigations in Britain, there were effects ranging from strong, significant and positive through to non-significant and neutral, to relatively small but consistent negative impacts of ILS on outcomes (Wood et al., 1999). One major reason is likely to be the different outcome measures used to assess impact.

The most consistent finding appears to be evidence of gains in certain areas of mathematics, particularly those relating to basic skills. This factor, together with the lack of generalisation of progress on the ILS system to external measures, suggests that what the ILS curriculum offers may be poorly aligned with what is demanded by the curriculum. This appears to become more marked with increasing age as the curriculum moves from an emphasis on basic skills to requiring ‘semiotic’ work.

In contrast to the findings on learning outcomes, there is far more consistency in findings regarding attitudes and behaviour (Hativa, 1989; Wood et al., 1999). These are not presented in this report as there is no evidence of a relationship between these and learning outcomes (Wood et al., 1999). As Wood and colleagues note, “One implication of this is that any exclusive reliance on "user satisfaction" as an index of the effectiveness of technology should not be taken at face value and <should be> treated with considerable caution in the
absence of converging evidence for effects on performance and learning" (Wood et al., 1999, p. 95).

Although methodological issues loom large in a consideration of the studies, deficiencies in this area cannot provide an adequate explanation of this variability. What does seem to emerge is that evaluations, especially non-independent evaluations, should be treated with greater caution. The overall conclusion is that the level of effectiveness of educational technology is influenced by features of the software, the specific student population and also the level of access to technology and features of the educational context, particularly how the educator or educational body organises for its use. It may be that if variation in a school’s use contributes to differential impact, then less control over implementation is likely to lead to more variability. Generally, independent evaluators have less control than where vendors install and support implementation. This, however, does not necessarily explain the obtained level of effectiveness, but suggests reasons for variability.

2.8 Conclusions and discussion

What is one to make of these studies of CAL in general and ILS in particular? To recap, there are few studies that yield an effect size that could be considered more than moderate (.3) or above the average effect size for all educational interventions (.4).

There is a need to look at the empirical, methodological and conceptual implications of the findings as suggested by Wood, Underwood, and Avis (1999). The main implications, as these authors see them, are in terms of conceptualising what principles of system design make for the most productive system-learner-teacher interactions. These authors look at the findings (of ILS) in retrospect and ask whether they could have been anticipated, or even explained in terms of what is already known from theories of learning and instruction and from the relationship between the way knowledge is represented and the context to which that representation generalises. The thesis being advanced is that there is a need to reconceptualise the way software is designed. Other researchers (H.J. Becker, 1992) agree that to obtain more effective learning technologies, the design needs to reflect more social theories of learning and, in particular, the system, teacher, learner interaction triangle.

Wood and colleagues (1999) spend some time discussing both models of learning and computer based learning and knowledge representation. Numerous researchers have criticised the models of learning that CAL and ILS embody. For Pea (1993a, 1993b), the type of learning where facts and skills are acquired is less beneficial to a student than problem-
solving activity designed to foster higher order thinking skills. For Becker (1992b), the lack of social interaction or meaningful interaction with the materials is such that genuine learning is inhibited. Interestingly, the use of co-operative learning in concert with an ILS was shown to lead to better outcomes in maths than using the ILS in an individualised way (Brush, 1996). Mathematics educators criticise computer-assisted learning systems as not allowing for personal knowledge construction, and asking the child to use methods that may not come naturally. The also criticise the programs for not recognising the importance of multiple representations and multiple strategies (e.g. Schoenfeld, 1992; Secada, 1992).

The other issue concerns the idea of representation of knowledge. According to Wood, Underwood and Avis (1999) “The assessment of performance which underpins an ILS is that of a traditional kind. Since the models used to characterise learner knowledge are based on these assessments, they constrain the definition of tutorial goals and the planning of routes through the teaching materials. So, any limitations on assessments are likely to be reflected in the restricted range of learning goals and pathways that ILS can support” (pp. 102-103).

The fact that learners seem not to obtain a conceptual understanding of the procedural skills they may gain is because such goals as flexibility of reasoning or problem-solving skills with more than one system of signs, are not the goals of the designers of current ILS. “Either one accepts that such systems have no place in the classroom at all, or one accepts that they must be integrated alongside other teaching and learning practices if they are to work as a significant contributor to learning and understanding” (Wood et al., 1999, p. 104).
1.0 Introduction

It is difficult to integrate the findings with respect to this question of value. The word value is used to suggest something beyond simple effectiveness. Many evaluations did not consider value directly or, if they did, they addressed the broad question of value in different ways. Few consider cost relative to alternatives or take this analysis a step further and consider cost effectiveness. As Dyer (1992) notes, "Assessments of educational programs concern themselves almost exclusively with effectiveness, mentioning costs only in passing or ignoring them altogether" (cited in Inouye, Miller, & Fletcher, 1997). Therefore, the findings relating to value from various studies are presented and an effort made to draw some conclusions from them.

1.1 Relative effectiveness

Despite the fact that meta-analyses show gains for those in the CAI / CAL condition compared to those in a control condition, the results are variable and, more importantly, can be seen as ineffective relative to other innovations. Research work in the area of meta-analyses shows computer-assisted instruction to be less effective than other innovations. The effect size for CAL is calculated to be .31, while the average for other innovations is .4. Hattie (1999) presents a table illustrating the effect sizes for a large number of innovations (See Appendix 1). In the literacy area, peer tutoring could be considered an alternative candidate as an innovation designed to enhance outcomes. The effect size for peer tutoring is .50.

1.2 Relative costs

A cost benefit analysis of the West Virginia BS/CE program (Mann et al., 1999) showed that it was more cost effective in improving student achievement than either class size reduction (from 35 to 20 students) or increasing instructional time or a cross age tutoring program. This is at variance with earlier work (Levin, Glass, & Meister, 1987) that compared CAI generally, relative to tutoring and class size reduction and found CAI to be more costly. There are a number of factors to consider in both accepting either of these findings and explaining the differences. The first is that the class size literature suggests that there is no appreciable effect until the class size is reduced to 15 (Glass & Smith, 1978).
Another consideration is the relative, and ongoing, decrease in cost for computer hardware. These factors make it very difficult to accept an analysis other than as being an indicator at one point in time.

As a part of the New York school system study of an ILS discussed above, there was an evaluation of system cost and a comparative analysis with a variety of current approaches to school improvement. These included a textbook based program (DISTAR), a computer-based program (IBM's Writing to Read), a school reorganisation and tutorial program (Success for All) and an intensive one on one reading program (Reading Recovery). These programs supposedly fitted the all-important criteria of comparability in implementation across schools as they had clearly prescribed requirements. Also, cost data were available. However, the meta-evaluation of the New York study rightly points out that the contexts where the alternative intervention programs were applied were undoubtedly very different which limits what can be said about their comparative costs.

The analyses focused on incremental costs, that is costs required in addition to those already provided by the school or district. These costs constitute the net present value of implementation over an eight year period and are presented as per pupil costs for those participating. Overall, although the resulting analysis produced variable costs per pupil for the ILS (depending on lab size), and Success for All (depending on grade level), the incremental per student costs of those ILS where a figure was able to be calculated were approximately twice that of DISTAR but half that of Writing to Read and much less than Success for All or Reading Recovery (the latter was over $5000 per student compared to the $200-$250 for ILS).

But, the authors of the meta-evaluation of the New York Evaluation study suggest that the analysis overlooks a key issue. What does it cost in additional per student expenditure across all students and grade levels for a given period of time to get the students most in need up to acceptable levels? This involves calculating the duration of incremental costs for each student, that is in relation not just to the particular students involved but the school's total instructional budget. It is pointed out that this is particularly important for a program like Reading Recovery since for a large, relatively short term investment the program is designed to move targeted students to a level where they no longer require remediation. Such an analysis would be improved with the addition of effectiveness measures as cost utility is meaningless if the intervention has no or limited effects.
1.3 Alternative programs and costs

In a report "Reading programs that work", John Schacter (1999) compares a range of school-wide reading programs, reading technologies, reading tutoring programs, and early child reading curricula and programs. He uses existing evaluations of the widely used programmes. In each case, Schacter describes the program; looks at the reported effectiveness; considers the professional development associated with the program, and its cost. He slates the quality of evaluations of various reading technologies saying that this makes it extremely difficult to determine if the technology is effective. He states that because of the lack of quality research, the standards for including a technology for review were lowered relative to other programs considered in the report. He cautions that even the effectiveness measures included in his report "should be interpreted with scepticism" (p.19).

Of the multitude of software available for reading, according to Schacter (1999) only 11 producers have tested their software's effectiveness against other methods of raising reading achievement. As Labbo & Reinking (1999) state, technology has been beset by this horserace mentality and by this horseless carriage rationale where the new innovation has to establish its superiority to an existing method. These authors, while not dismissing the need for such research, point out that it has limitations. However, the conclusion Schacter reaches is that from the research, few early reading technologies have proven to be effective with few producing substantial gains in student performance. He says that most reading technologies are expensive and the return in terms of achievement is "modest at best" (J. Schacter, 1999, p. 19).

He compares the cost and effectiveness of a number of reading programs. Tables summarising Schacter's findings are included in Appendix 3. In contrast to the school-wide reading programs like Success for All, Exemplary Center for Reading Instruction, Junior Great Books and Project READ which all have the top rating of comprehensive evidence, meaning that a large amount of research has been conducted with the majority of findings showing reading improvement, the reading technologies have no such ratings. Most have the rating little evidence: research with mixed results. SuccessMaker has a rating between the category "little evidence" and the category "evidence: some research with majority of findings showing reading improvement". Schacter bases this classification on Kulik's meta-analysis where 20 evaluation studies used components of SuccessMaker with an average effect size of .4; on Becker's reanalysed results reporting that independent studies reported modest or negligible effects, and a recent study by Shapley (1997) that found that SuccessMaker had no effect on reading achievement. Some of the reading technologies like
the Waterford Early Reading Program and Breakthrough to Literacy are expensive relative to
Success for All, the most expensive of the school-wide programs ($600-$900 per student
against a maximum of $150 a student. Note: Cost for SuccessMaker given as "highly
variable").

In Schacter's (1999) analysis, tutoring programs fared better as regards evidence of
effectiveness than technology driven reading programs.(It should be noted that Schacter did
not look at Reading Recovery). He considered both professional and volunteer efforts. The
Howard Street Tutoring Program, a volunteer program, was seen as highly effective and
relatively inexpensive at $80 per student. It is a small, community-based program and the
study of its effectiveness included only 50 second and third graders in inner Chicago who
were having reading difficulties. They received tutoring twice a week for one hour a time
while the control group received no tutoring. The treatment group performed much better
than the control on word recognition and spelling and substantially better on oral reading
(effect size 1.07).

Book Buddies is a similar program for first graders (Invernizzi, Juel, & Rosemary,
1997) and again there was a large effect in terms of word recognition measures when those
that attended more tutoring sessions were compared with those who did not (effect size
1.29). Note, that there was no control group and there may have been systematic bias in that
those who attended more sessions may have had certain characteristics that predisposed
them to perform better. The cost for Book Buddies is the manual ($20 a copy per tutor) and
the time of a graduate student to co-ordinate the program. More expensive is Reading One-
One which uses trained and managed para professionals. Here the cost is less than $10 a
student per tutoring session with training costs $10,000 for three schools. Research (Farkas
& Vicknair, 1996) showed the program of three to five, one hour tutoring sessions per child
over a four to six month period to raise the grade equivalent score by half a year. There were
no comparison groups only pre and post-test data.

The most expensive tutoring program in Schacter's review was Early Steps, which
uses trained teachers to work one on one with students at risk (bottom 20%) for reading
failure. In the one study of its effectiveness, 23 matched first grade pairs took tests in which
the treatment group did much better (in word recognition effect size .47; spelling effect size
.80; passage reading .77). Both groups received equal instruction time and tutoring time.
While the treatment group was tutored with Early Steps, the control received "generic
methods". At the end of first grade 52% of the treatment were at or above grade level while
only 23% of the control were. Of the tutoring programs classified as having evaluations
showing comprehensive evidence (ie. large amount of research conducted with the majority of findings showing improvement) or evidence (ie. some research with majority showing improvement), the costs per student range from $20 to $833.

From Schacter's analysis, it is clear that reading technologies are relatively expensive and largely unproven when compared to either school-wide reading programs or tutoring programs. It could be argued, however, that reading technology programs are difficult to evaluate as they are a moving target and also that they are able to provide other value-added components. Some of these other components are alluded to in data collected as part of the U.K. evaluations of ILS.

1.4 Non-cost, non-effectiveness related value of ILS

The U.K. evaluations of ILS asked for ILS users' assessments of value for money, using interviews and questionnaires administered to teachers and senior management. The teams working on the evaluation noted that the adoption of ILS in a school had system-wide effects that could influence both the benefits and costs associated with using the technology. The qualitative equation they identified was Value = learning + attitudes + staff development + enhanced learning environment + IT image <divided by> hardware/software/maintenance costs + staff training costs + disruption.

It is clear that, potentially, ILS could provide a shared focus for group thinking about the use of IT in relation to a school's policy in the area of technology. It is a practical way of exploiting IT for curriculum delivery and can be a way of enhancing a school's IT image within the wider community as well as providing a shared resource for use of the community. In addition, on the benefits side, the use of an ILS demands the development of new knowledge and practices by teachers like how to integrate the ILS into the curriculum and with their other teaching practices or how to interpret and use system feedback. The perceived benefits reported include increased staff confidence in IT systems and a contribution to professional development by raising issues fundamental to teaching and learning. However, although an ILS may enhance IT skills, it is a costly and laborious way of doing so. What emerged from this consideration of factors that may add value was the sense that even after two or three years of experience with the system, schools were still adapting to and experimenting with the demands of integrating ILS.

In terms of the possible costs, it was estimated that set up costs in one secondary school were approximately equivalent to two junior members of staff for a year. There are no
figures for estimated costs offered in this instance or in the U.K. study as a whole. In any case, schools varied in the extent of the IT infrastructure that they acquired to support ILS and they purchased systems with different direct costs. Plus, the real cost to most of the schools was less than the commercial going rate. Thus there is no obvious common denominator for a single or simple cost benefit analysis.
Part C: Preconditions for Effective Use

1.0 Introduction

The findings of the first large scale, comprehensive, longitudinal study of the impact of information communications technology on student achievement across age and curriculum areas, concludes that “in particular circumstances”, IT had a highly positive impact on children’s achievements (Johnson, Cox, & Watson, 1994). The conditions and circumstances are the focus of this part of the report. Particular attention is given to the context of use of ILS. The research literature is presented first, then the second section presents and discusses data gathered from New Zealand schools where an ILS is currently used. The aim is to consider the factors that affect the learning outcomes of students.

There are several issues that arise from the literature that concern effective use. The first is a group of considerations that could be termed organisational. The second could broadly be termed pedagogical. Under this broad rubric comes the theoretical stance of software and allied issues of individualisation and match with school curriculum and assessment and teacher’s preferred methods of teaching.

1.1 Organisational aspects

As discussed previously, research literature on ILS provides explanations for their variable results in terms of successful outcomes in schools. The success or otherwise of ILS is highly context dependent. While some of the variation observed in outcomes is due to student and school populations, another potentially more powerful source of variation stems from aspects of implementation. The same ILS operated differently can produce varying outcomes. As Smith and Ragan (1993) note “In drawing the line of causation from the instruction to the results, it is critical to be able to identify the degree to which the description of the program represents what actually occurred during instruction with the new program” (p. 416).

Two researchers who have done extensive research on ILS have developed a set of factors that facilitate or impede implementation of ILS. They suggest that results from ILS have often been poor since the implementation is done poorly and may even run counter to the implementation designed by those who provided the system. They found that many schools are not aware of the implementation models provided with the ILS (Van Dusen & Worthen, 1993). They suggest four distinct components of implementation that appear to be
essential if ILS is to have its desired effect on student achievement. These factors include:
(1) adequate time on the system; (2) teacher involvement; (3) curriculum integration, and (4)
staff development. As it appears that there is no guarantee that a system that has worked 
well in one school will do so in another, schools need to be aware of the likely conditions that 
affect outcomes.

A factor that has emerged is the support from management and the incorporation of 
the ILS into the school mission or plan. The incorporation of the ILS may need new ways of 
working for teachers so training is required. As the British study notes, staff training must be 
an integral part of a school's long-term implementation strategy. It is not sufficient to send 
one staff member for training and expect a cascade effect. Simply expecting new staff to 
absorb the culture as they work alongside experienced colleagues is not enough (BECTA & 
McFarlane, 1999). Like teacher's use of technology in general, there are stages of 
participation in the implementation of ILS. The teacher may move from the non-participatory 
level where the teacher "drops off" a class at the lab, through to a higher level (termed 
integrator) where a teacher is able to manipulate the ILS sequence so that it better matches 
classroom instruction. The final level is that of extender where the teacher has fully 
integrated the ILS and classroom curricula (Clariana, 1992). It is suggested that even where 
teachers are competent IT users, an ILS presents new ways of working (BECTA & 
McFarlane, 1999). Where teachers do not use the ILS systems as intended, this leads to 
possible underestimation of the effects (Van Dusen & Worthen, 1994; Worthen, Van Dusen, 
& Sailor, 1994). The training has to be ongoing, according to teacher users, and has to take 
account of the differentiated roles from system manager to tutor (Lawson, Underwood, 
Cavendish, & Dowling, 1997).

There are a number of models of implementation of ILS, both for staffing and for 
organisation. Each model has its advantages and disadvantages. The most common staffing 
models (and all are to be seen in New Zealand schools) are to have the teacher coordinator 
static in ILS room, supervising most or all ILS classes; to have the teacher and students 
move to the specialist ILS room; where the teacher works in his/her own room; to have part 
of the class move to the lab with an ancillary staff person, or to have the ILS as self-standing 
with supervision either intermittent or indirect. Each raises issues but a central one is the 
extent of home or subject teacher familiarity with what the students are doing on the ILS so 
that they can integrate both experiences to maximise learning gains. Tables of the models 
and their advantages and disadvantages are adapted from a guide to good practice (BECTA 
& McFarlane, 1999, p.11) and presented in Appendix 4.
As for models of organisation, these range from a room set up for just ILS to a room that organised for half the students to be on the system and half off, to a distributed network, to a pod, and to a stand alone machine. Clearly, the best implementation model varies with ILS system and with school configuration. However, from the British experience, it seems that sending students to another room for their ILS work is proving less successful because the ILS experience becomes disassociated from the teacher and work in the classroom.

There are some factors that have been shown to be crucial in all models of implementation. Van Dusen & Worthen (1994) suggest three aspects of implementation to be positively related to outcomes. These include the average weekly amount of time on the system; the average number of lessons completed and the level of integration of ILS work with classroom instruction. For example, there has been shown to be a direct relationship between time on the system and gains for low SES students and increased intensity benefited low achievers more (Hativa & Becker, 1994). There is a large difference in the times recommended by suppliers of the various ILS products. From the British evaluations, it is clear that exceeding these times or aggregating into fewer, longer sessions could lead to gains being lost or reversed.

1.2 Pedagogical

Theories of learners and learning have undergone radical change during the past 15 years and this has important implications both for how educational technology applications are used and assessed and for their place in the wider pedagogical context of learning. In many cases it is the learning approach that determines the success or failure of the technology intervention (Means & Olson, 1995; Seeley, 1995). The theoretical context affects assumptions about curriculum, instruction and assessment. Further, there are claims that successful implementation of CAL depends on how well the learning theories adopted by teachers to guide their classroom practice are consistent with the learning theories adopted by program developers that guide the design of the educational software to be used in classrooms (Cognition and Technology Group at Vanderbilt, 1996). Computer-assisted learning of the type represented by ILS is seen to neglect the socially based nature of learning; to not suit all learning types, and to present a disaggregated and hierarchical form of learning that does not readily generalise.

One of the arguments for ILS implementations having a disappointingly small effect has to do with "the ideology of autonomous tutoring by software and individualised pacing that accompanies their marketing and operational structure" (Becker, 1992, p.6). This
approach emphasises the interaction between the individual and the content, not the more recent view of learning occurring in the social interaction between people. The developers of ILS, according to Becker, need to base their products on a "more complete and accurate theory of effective classroom instruction and learning" (Becker, 1992, p. 7). It does seem that exploratory classroom environments with technology like CSILE (Scardamalia & Bereiter, 1996) and Anchored Instruction and other designs of the Cognition and Technology Group at Vanderbilt (Cognition and Technology Group at Vanderbilt, 1996) show successful outcomes. Learning tasks are team based, often interdisciplinary, oriented towards the solution of complex, real-life problems and utilise a variety of technological means. The nature of the learning outcomes expected from these exploratory environments may be different from traditional goals although there is evidence of enhanced performance also on traditional tasks (Scardamalia & Bereiter, 1996).

It could be argued that ILS software by providing explanations at the appropriate level, by diagnosing learning difficulties and providing feedback, is making learning more likely. The fact that there are results suggesting that an ILS favours the high and low achievers, suggests that its greatest advantage lies in establishing an appropriate level of instruction. But, Becker (1992b) argues that ILS, in making sure that instructional level is appropriate, neglect instructional quality. Drawbacks to individualised learning centre around the lack of social interaction in learning, although the research has often discussed drawbacks in affective rather than cognitive terms.

Long-term, individualised use of ILS, which is generally the recommended method for using these systems in schools, can have a variety of adverse effects on students (Becker, 1993; Brush, 1997; Mevarech, 1994). Long-term usage is connected to anxiety and hostility towards subject matter (Brush, 1997; Lepper, 1985), increased feelings of inadequacy and helplessness (Hativa, Swissa, & Lesgold, 1992; Mevarech, 1994), and a general dislike of ILS activities, particularly amongst low achieving students (Brush, 1997; Hativa, 1994), along with a decrease in teacher-student interaction (Becker, 1994). There is a suggestion that the ILS instructional model could be adapted to overcome what might be seen as the drawbacks of individualisation. If the ILS were more flexible, it could be structured to support mixed ability group work or a mixture of individual and whole class work.

Another pedagogical criticism of ILS systems is that they have sequential structure whereas not all learning occurs in this way. CAI, as an instructional model, may not be suitable for all learners. Ross and Shulz (1999) found that abstract, random learners may be at risk for performing poorly with certain forms of CAI. Other research has shown that the
quality of learning material is enhanced if the material is designed to take account of learners’ individual learning styles (Rasmussen, 1998; Riding & Grimely, 1999).

The way an ILS conceives of knowledge has implications for planning classroom activities. The term integration in integrated learning systems is seen by some as a misnomer because, as far as learning is concerned, the most important factor is taking the learning from the system and applying it in the wider world. There is evidence that, because of the form the learning takes, generalisation does not readily occur. There should be more emphasis in the ILS work on integrating different skills and applying them in a real context. Currently the teacher has to ensure this happens.

There are implications of these features of the theoretical approach of the software for planning of sequences of classroom work. Basically educators need to consider which areas of the curriculum that an ILS can usefully provide instructional content and computer activities for individuals and which areas are for teacher-led whole class learning. There are skills and competencies that are hard to master directly from computer tutorials and others that are not part of an ordered or hierarchical set. ILS developers, according to Becker (1992b), could incorporate both teacher presented lessons and material to be used in off line sessions as well as a management system that allows students to start at the same level, then branch to other related or unrelated enrichment type activities.

Teachers have an important role in creating an effective CAI learning environment and in ensuring that learning tasks align with clearly set learning goals and monitoring student progress to make sure the CAI implementation is leading towards the goals. Evidence points to the teacher's vital role in the success of ILS by directing student learning and effort (Hatvia & Becker, 1994). Where ILS sessions are supervised by someone who does not offer support, achievement outcomes, particularly of lower achievers, suffer (Becker, 1994; Clariana, 1990; Taylor, 1990). Teachers need to encourage students while on the system; monitor progress and provide class work that inter-relates.

However, a point to note is that many educational program vendors (e.g. ILS vendors) use ease of implementation and availability of courseware, training, and support from one source as major reasons for buying their products. These perceived benefits are not without critics. Studies have shown that, in many instances, a belief in the turnkey approach leads to lack of teacher involvement during CAI sessions (Brush, 1998). Becker (1994) stated, "Because such programs can run with little intervention from the teacher … it is tempting for schools allow ILS programs to run essentially unattended except for the technical support
provided by system managers …" (p.78). This lack of teacher involvement has led to improper coordination between classroom-based and computer-based instructional activities, inadequate student support while they are completing ILS instruction, and lack of teacher understanding regarding effective strategies and procedures for using ILS (Becker, 1993, 1994; Brush, 1997; Hativa, 1994; Sherry, 1990).

Effective integration of computers and other technology, according to McGrath (1998), requires that teachers (1) become comfortable with the technology itself; (2) explore software, CD-ROM, Internet-based and other curriculum resources to identify those that might enhance and enrich their current curriculum; (3) review their curriculum to determine how best to integrate these technology resources into their lesson plans' (4) revise the lesson plans to incorporate the technology resources; (5) experiment with the lessons in the classroom; (6) assess how well things worked; and (7) refine the lesson.

The manipulation of the ILS by teachers to ensure good curriculum matching is necessary for students to gain maximum benefit (Blickhan, 1992; Clariana, 1992; Schnitz & Azbell, 1990; Shore & Johnson, 1992). The curriculum content match and the level of integration of the use of ILS have been shown to be major factors affecting learning outcomes (Becker, 1994; Clariana, 1992; Hativa, 1994; Shore & Johnson, 1992). If the goal is to assist schools to achieve a particular target in literacy or numeracy, measured by a test (eg. in U.K. SAT; in N.Z. PAT, and in the U.S. a state-wide mandated test), then there needs to be a match of both the curriculum and proposed methodologies. Further, the ILS needs to reflect the task types encapsulated in the assessment test.

Wood, Underwood and Avis (1999) claim that demands on teachers in ILS work are greater than those in individualised work on single topic computer programs as the diversity in the range of potential problems is huge. Becker (1992) argues it may be more difficult for the teacher because of the difficulty of attending simultaneously to students who are working on a wide range of tasks and skills. Adjustment of the level of description of learning material to match the moment-to-moment needs of the individual learner is of primary importance in this kind of tutorial environment (Mayes & Fowler, 1999).

At first, in the early period of implementation, teachers feel that there is a reduction in workload as the computer marks and selects further, appropriate student work. However, this is short lived as teachers begin to appreciate the need to produce differentiated programs to blend in non-ILS lessons. There is help needed for students with limited comprehension and limited self-learning strategies. To integrate ILS into their classrooms
Teachers need to work at the integration of what might be conflicting methods—whole class instruction and individualised teaching in the ILS work. They need to understand the ILS curriculum, its content, context and methodology thoroughly in order to relate experiences on ILS to other contexts and related skills and knowledge. This latter point is all-important. The findings from the various evaluations of ILS suggest that it is useful for teaching procedural skills, where practice and repetition are important. It is the teacher who has to make the link from such skills to real world application. As the U.K. evaluations note, the "Is this an ‘add’, miss?" phenomena is prevalent when ILS are used. Students improve in their mastery of basic skills, but they still do not know how to apply these skills.

Teachers have to work differently in terms of managing learning both monitoring task behaviours and progress (Parr, 1997). Time and training are required to read and analyse the system's diagnostic reports, as some can be quite complex. This varies by system. Some simply summarise the number of right and wrong answers while others provide numerous, in one case 60, standard reports. The U.K. studies found that data were used more effectively when selections were made about which elements of performance to use, instead of attempting to use everything the management system could generate. Teachers needed to establish that recurring patterns were of significance, not individual instances in making decisions about a course of action for individuals.

They also need to monitor closely to ensure that the resource is being used to best effect. Given that the results produced by the ILS systems are both highly varied and of dubious external validity, it is important that schools have in place a parallel process to monitor student progress. Schools also need to ensure that the process is geared to the curriculum and assessment procedures that currently operate at classroom and school level. The U.K. evaluators note that it is common for teachers to be enthusiastic about the system, especially initially, and to be adamant that students are benefiting, even in the face of independent evidence that students are, in fact, not making significant progress. One factor is that teachers can be influenced by student's positive behaviour and attitudes and fail to distinguish this from improved performance. Also, evaluation should be ongoing as responses to ILS may change over time or the characteristics of a new cohort may differ from those of an earlier cohort.

There is evidence from both the U.K. studies and the use in the U.S (McFarlane, 1994) that use of ILS to identify the range of ability levels, then to group children with common needs and plan material for the week does much to allay concern that teachers may have about either the considerable age range (e.g. in small rural schools) or the wide ability range.
evident in some of the large urban schools. To help the teacher create small groups of students with similar needs so that lessons and computer activities could be structured for these groups, the reporting system of current ILS may need to be modified as it does not present reports that help teacher led remediation; that is reports are not organised by skill or lesson or unit so that students who had found difficulty could be identified.

Teachers are a key factor in implementation. At the teacher level it has been shown that even within a school district using the same system, it was implemented differently. Integration with classroom structures accounted for 56 percent of the variability in the degree of ILS implementation relating to outcomes.

As far as pedagogical conditions to maximise outcomes from use of ILS, Becker (1992) says that developers of ILS are "deceiving themselves if they believe that their systems, as currently designed, will lead to long-term consumer satisfaction by administrators, teachers, parents and students" (p.14). Present set ups ignore the social nature of learning and the requirements of group-based instruction. Greater diversity in instructional organisation is needed as is some investment in helping teachers use the resource along with their own talents and other resources to help develop their students.

2.0 The New Zealand Experience: Views of Schools

2.1 Context

The Ministry of Education established a special pool of money to which low decile (1 and 2) schools could apply in order to instigate projects that aimed to raise the literacy and numeracy levels amongst their students. In the first funding round, 17 schools were funded, while in the second round, one was funded. These funds were for one or more stations operating SuccessMaker software.

2.2 Participants and procedure

The Ministry of Education provided details of schools that had been funded to purchase SuccessMaker. The Research Division of the Ministry also wrote to these schools explaining the project. Schools were contacted and invited to participate in an informal discussion type interview. In this way, views were obtained from 14 of the 17 schools funded in the first round of the grants. Thus, these schools had been using SuccessMaker for about a year. Interviews were with key personnel associated with the implementation and operation of SuccessMaker. In most instances, this was the principal but, in some cases, it
was the principal and the person responsible for co-ordination of SuccessMaker while, in other cases, it was only the latter person.

Informal interviews were arranged at a time convenient to the school personnel involved. All but two interviews were by telephone and lasted, on average, half an hour. Two site visits were also conducted. The same areas were broached in each discussion. These areas of discussion included how the school became interested in the program; details of how the school organised to use the program; which children were selected and how; what the assessment was of how the program was going, specifically in terms of enhanced learning outcomes; what data had been used on which to base a judgment, and what factors were seen as important in ensuring that the implementation worked as well as possible. In addition, two schools volunteered copies of internal evaluations of SuccessMaker in their school.

2.3 Issues from the discussions that suggest preconditions likely to be associated with more effective outcomes

2.3.1 The idea of one tool among many

There was a pervasive view that this form of innovation was simply another tool to employ in learning, not one to be used to the exclusion of others. Some schools used the term “eclectic approach” to describe the fact that they had tried several things to try to enhance learning gains in their students. The feeling was that there were so many needs in these low decile schools that teachers and management were prepared to try different approaches. Although one described it as a powerful tool while another said it was the best use of resources other than a teacher that s/he had seen, yet another stressed that an effective classroom teacher is THE best resource. One felt this tool offered something very important, a “screen”, that was something “kids are readily hooked in to”. Another saw the use of SuccessMaker as a way to diversify IT learning.

2.3.2 More than just the SuccessMaker software is needed

The point was made strongly by more than one respondent that the software needed good hardware to run effectively, including a server. The cost of the software was only part of the outlay, albeit a considerable one. Another comment was to the effect that it is important to create a comfortable learning situation so that details like appropriate chairs and good quality headphones with long enough leads were important. However, the overwhelming message was that personnel were necessary to ensure optimum use of the
software and this often entailed additional cost or deploying of people currently used elsewhere.

2.3.3 Personnel are central

Most interviewed stressed that SuccessMaker is not a stand alone intervention and they thought that it could not be maximally effective without both skilled oversight of the project and without personal assistance being available to students as they worked on SuccessMaker. Some, however, did consider that the tuition given was sufficiently good that SuccessMaker could work with an aide present and a teacher in a management and overseeing capacity. One principal made the point that most children do not need a teacher’s help to make basic progress as SuccessMaker has material sequenced in “bite-sized pieces” but they do need help to place that basic knowledge in a broader context.

Trained personnel are necessary to place children accurately and to monitor their progress, making the necessary adjustments. One comment concerned the fact that it is necessary often to make what seem like small adjustments for example to the default time available to read instructions. Others commented that it requires someone there to give assistance or tutoring, to observe when students have a problem and “hop in, suggesting, for example, that they re-read or scroll back up to find a key idea etc”. One principal felt that a “tutor” was definitely required, one with “belief, passion and know-how”. A teacher with responsibility for SuccessMaker said that s/he “observes, prompts and teaches” all the time while the children are working on SuccessMaker. One school observed that for younger children someone is often needed to read the material in maths as the reading level required in the maths activities seemed to be beyond their current reading ability.

There were a few schools that ran the program without a person dedicated to interact with the children working on SuccessMaker. This type of set up was often in a small school where the principal assumed the role of keeping an eye on the child on the computer, or in a school where there may have been an aide or teacher either in the same room working with others or nearby. In these cases, the child worked independently but help was available. In one such set up, the principal reported that SuccessMaker is “independent and just ticks away with little input required”.

The view of most was that an experienced teacher should run or at least oversee the operation of SuccessMaker. “It is probably necessary to have someone responsible for SuccessMaker” and ideally that someone should not to be wearing too many hats already!
This person had to “know SuccessMaker really well” and have “vision and breadth to see how it could be integrated”, in order to consult with teachers about where the program might fit in units of work. Other than a brief for integrating SuccessMaker into the curriculum or for monitoring and helping students while they were engaged on SuccessMaker activities, there is also a need for someone to run the management side. This includes setting the system up, getting SuccessMaker running and interacting with the software vendors and hardware suppliers. This is not always a straightforward task to network client computers or perhaps re-format due to hardware or software problems.

Management was also seen to include the collating and interpretation of the diagnostic information that the system produces. One principal felt that teachers become increasingly interested when they see the results. Another school felt that an important factor in getting SuccessMaker to work optimally was to have someone to get the data out on a weekly basis. This person had to be trained to be able to explain the data to the classroom teacher and to help link what they are doing on SuccessMaker to the classroom activities. In this way, it was felt that teachers would have more ownership and see it less as “something done to kids in their class”.

Further, comments suggested that there needed to be someone, “a teacher who commands respect” to help get staff on board, particularly in cases where staff have found it “difficult to accept”. This expert, up-skilled person can “expose other staff to SuccessMaker and conduct workshops”. As mentioned above, the person needed to liaise with staff to see where the program fits with the classroom.

2.3.4 Importance of selection, placement and tailoring to needs

Most of the schools began by selecting the children most in need of help with literacy or numeracy to work on SuccessMaker. In one school’s case, this meant that a third of the children on the program were regarded as having “special needs”. Selection was done carefully, employing appropriate measures like normed tests (eg. Before 9, PAT, Schonell, Burt), together with teacher recommendation to decide who was to be placed on the program and approximately where to start each child. Considerations were sometimes whether the child could operate relatively independently in order to read instructions or keep on task.
The program was seen by a number of schools as only “working for some”. A view was expressed that the very poor did not benefit as they often did not have skills to work independently and required one to one with a tutor. “It is not designed for the child with learning difficulties who needs to be stepped through a problem”. The feeling from another interviewee was that it was helpful for those who had reached a plateau behind the average. However, based on experience this level should not be too far behind. One school had started with children four years behind expectation and found SuccessMaker did not work for them. Even with those two years behind, this school commented, it can be problematic. The data used to make this judgment came from standardised pre and post-tests, not from the program-generated data. This finding, that those considerably below average do not benefit, was supported by two other interviewees.

It was suggested that SuccessMaker was a different style of learning and that some children benefited enormously while others did not. There was a view expressed by one school that the style of learning suited some but many missed the social interaction. For some schools a selection consideration was that it was important to have children who are able to be self-motivated. One school, in particular, noted that many children did not find SuccessMaker intrinsically motivating after the initial novelty of working on a computer wore off. They found that students tired after 5 to 6 weeks on the program. They instituted various measures to try to hold interest and raise motivation. In the second intake of students to work on SuccessMaker, they attempted to include those who had demonstrated more independence and self-motivation but the outcomes were still disappointing.

Initial placement was seen as being very important as the accuracy of it determined whether the child would experience success or not. One school reported that they did not really understand SuccessMaker when they got it, thinking that it would generate the placements and sort children into levels. They found that it did not work like that and it seemed to be more effective if the teacher had input into checking the levels and the standard the children are working at. For example, using the computer-generated placements led them to set too high a standard and so the children were not experiencing the early success so important to motivating them. Several schools acknowledged that placement is “tricky”, partly because it is often difficult to complete the number of sessions recommended for accurate placement for some children with absenteeism etc. There needs to be good testing and skilled observation to pick up when a child is not succeeding and needs a change in level. One interviewee commented that it can “take ages to notice when a student needs a change of level because they are not succeeding”. The point was also
made that the question of correct placement can affect the statistics the system generates so when one school finally decided on the placement of a child, they re-enrolled the child.

2.3.5 Time and knowledge to utilise, assimilate, unpack

“There is a great deal of time needed to be put in to ensure that the most is gained from this software”. There was expressed on several occasions the notion that there was much more in this program than the school realised when they first acquired it. Nearly all said that it takes time to become familiar with what is in the program, more time than most schools had been able to put into it to date. The second round of training provided by the vendor was seen as being very helpful in pointing the way to enhanced use of the program, particularly in relation to the diagnostic data generated.

Most schools currently had only one or two teachers “up-skilled” so that they could operate the SuccessMaker software. Teachers not directly involved were described variously as “compliant”, “supportive”, “positively accepting” and “disinterested”. One larger school had five staff involved and several other schools recognised the need to involve more staff.

As knowledge grew, schools saw themselves as able to make more flexible use of the resource. For example, at the beginning, they often grouped children with a common need so that they could all work on a particular unit of work at a similar level. Now most feel that with more experience and training, they can support each one on a different, individualised program. There was also the comment that time is needed to reflect and review.

2.3.6 Lack of hard data on which to base judgements

There was generally a positive response to SuccessMaker from the people interviewed in all but isolated cases. One principal “pleasantly amazed” at the progress children have made and another talked of the “positive feedback” received about progress of the children. Yet another felt it had had an impact on children at their point of need. One school was quite upfront and honest in admitting that they had not yet done the “homework” necessary to accurately monitor any gains.

However, it is notable that one of the schools that were less enthusiastic had taken a hard look at progress by employing standardised tests, external to the program. For this school, the children selected because they were two to four years behind average in literacy
or numeracy, did not make demonstrate learning outcomes from working on SuccessMaker. This school wanted to base their evaluation of worth of SuccessMaker on measurable data, saying that attitude was “not worth looking at”.

However, few schools considered the use of data external to the program. One school took measures after each 10 weeks on the program, namely a Burt and running record but had selected the students to participate partly on PAT so there was no direct comparison in order to gauge progress. Another school reported using the Burt and running records to evaluate SuccessMaker.

The data the majority of schools used to inform their judgment about the success of SuccessMaker largely came from two sources. One was informal observation, either the personal observation of the key person connected with the program or teacher reports to this person. The other source was from the gains data (months per hour on program) or other data (eg. percentage correct or diagnostic data) that the program generated. Not all schools were yet equipped to extract this information. Only one school questioned, in the course of discussion, the real meaning of the data that was generated by the system. They questioned how SuccessMaker worked out the gains; how accurate these were and how they related to the curriculum.

Most schools relied on anecdotal data for formative evaluation. Interviewees reported teachers as being “pleased with progress” or that teachers had, for example, commented on the fact that children have learned a particular instance like the decimal point or keyboard skills. “They <the classroom teachers> are definitely seeing an improvement in maths, perhaps from the extra mileage”.

Some observations concerned work habits such as willingness to tackle work or being on task and knowing routines and these were favourably commented on. “The children as developing good routines; they know the timetable and what to do”. The program was seen as a way to “concentrate the concentration”. The observation was that children have focus when they work on SuccessMaker.

Many of the comments reported concern the affective dimension of learning. Children were seen to have a positive attitude to SuccessMaker although one school commented that the juniors were keen but the seniors less so. Children were generally seen as motivated to take part in SuccessMaker time slots. They “were really aware of what they are there for, namely, to learn”. Children “hunger for it” said one. They were a “captive audience" for this
type of learning. Another described the children as excited by it. Other affective gains were seen to be in relation to self-esteem or confidence. One principal said that the greatest gains were in this area.

2.3.7 Divided opinion over whether to integrate or not.

Although schools did not generally see SuccessMaker as stand alone in terms of the student operating independently and alone, many saw it as a stand alone in terms of the curriculum and classroom program. “The links are not there to the curriculum. It is a stand alone thing.” One said it was not built on the classroom program as it was skills rather than theme based. Most interviewed did not raise this issue of match or integration as a concern but saw the function of SuccessMaker as providing the “very basic, mainline stuff”. Others saw it as being related in that it was based on the curriculum areas.

In one or two cases, the school believed that the program should be integrated. In these cases schools aim to get a wider range of staff involved to see what SuccessMaker can offer specifically in terms of the curriculum. One school, in particular was trialing a classroom model of implementation to attempt to integrate the material better and to ensure that the classroom teacher knows what is happening. Another, when using SuccessMaker with the older children, is linking it into class work so that the children use it during some of maths class time and work on similar material to that covered in class. In another effort to forge links, teachers who oversee the program take reports or the results of diagnostic reports back to the classroom teacher and discuss them to enable the teacher to make links to the classroom work.

2.3.8 Organisational systems

One principal commented that the reason s/he considered that SuccessMaker worked was that the school had organisational strategies in place. Each of the schools interviewed appeared to have had to devise systems that allowed the maximum number of children access to the resource in blocks of time that they considered most efficacious. They have had to develop a way of facilitating the learning while on the system and a way of dealing with outputs.

With respect to organising for use, generally, 30 minute time slots were selected, as recommended by the vendors. Some schools, however, used only 15 minute time slots and others had some children working for 45 minutes (in one case the teacher in charge thought 30 minutes is actually enough). One principal commented that he wanted the children to stop
while they still wanted to continue and would, therefore, want to come back. Time on the program was sometimes adjusted both according to the type of work the children were engaged in (maths were in shorter sessions than literacy activities) and according to the age or ability of the child. Most schools operated on three sessions a week for each child but a few had two or had children working every day and, in one case, twice a day. The blocks of time that a group of students were rostered for SuccessMaker were mostly a term of ten weeks although one school used a Reading Recovery model and had children for a maximum of two terms and another used five week modules.

There needs to be systems for lots of specific communication, for example between the person responsible for SuccessMaker and the classroom teacher was the comment of one principal. The specifics of such systems, in most cases, had yet to be articulated.

### 2.4 Discussion

The largely positive response of the schools to SuccessMaker was not unexpected. The United Kingdom teachers were positive about ILS even when faced with contrary evidence about children’s progress. The explanation was that they based their perceptions on children’s positive behaviours and attitudes rather than on evidence of learning, particularly generalised learning. There is evidence that New Zealand educators are similarly basing their perceptions largely on affective and behavioural indicators, not systematically observed or recorded. To date, there had been little use of learning outcomes data other than limited use of the system generated data. For various reasons previously outlined in this report, this is unsatisfactory.

The emphasis placed on the role of the teacher in the ILS environment by the schools interviewed echoes that of the research literature where the teacher’s role is seen as vital in ensuring assistance, monitoring and class work that integrates. The latter aspect, manipulation of the ILS environment to attain a good curriculum match and integration with class work, was a current concern of only a small minority of schools. It may be partly that this requires extensive knowledge of the software and most schools were at capacity, using the available resources to get the program operational and to get students selected and working on the program. They had often yet to learn how to utilise some of the basic features of the diagnostic system like gains data, let alone acquire the level of expertise across staff necessary for integration.
Also, many were at the stage of using SuccessMaker for very basic skills work involving, for example, practice and repetition. When these basic skills are learnt, the literature suggests that teacher intervention is needed to ensure that children can apply them. However, it would seem that such integration and links to the classroom are necessary for effective use. Further, as the results of the U.K. (and one of the New Zealand) schools show, the ILS activities employed need to reflect those of the target assessment.
Part D: Reflection and Recommendations

The literature on computer-assisted learning does not provide a clear picture of the value of this form of learning. The analyses of quantitative measures of outcomes seem to imply that computer-assisted learning, on average, is no more effective than other types of interventions and may be less so. However, results with respect to enhanced learning outcomes are variable. This may be partly because of the nature of the research; partly because of the constantly changing and varied nature of computer-assisted learning but, more likely, this is because of widely varying contexts and the complex and interactive influences operating when computer-assisted learning is introduced into such contexts.

A major focus of this report was a discussion of the use of integrated learning systems in terms of enhanced literacy and numeracy outcomes for students. In this context, it is useful, by way of conclusion, to consider certain comments made by the ILS evaluation team from the United Kingdom. They reflected on what ILS can deliver in more general and abstract terms. However, the comments apply equally to many other forms of CAL.

The questions they pose fundamentally involve the nature of knowledge representation and the models of the learner and learning contained in the software. A major focus of the questions concerns the way that systems such as ILS both construct and evaluate models of learners' knowledge and the rules that are used to sequence and support learning. It is important to remember that the details of such models are not transparent. Although the basic theory on which some systems are based is published, the details of the way the theory is applied, in other words the algorithms that drive the system, are clearly not available (they may be seen to be commercially sensitive). This fact makes it difficult to comment in detail on the pedagogical strengths or weaknesses of the approach.

It is most important to consider the design of systems such as ILS because they are only one way in which technology can be used to support learning. In fact, as discussed previously in the report, more open-ended, problem solving use of computers in learning (such as CSILE) have consistently shown learning advantages. We constantly need to ask whether the type of learning approach offered is one we endorse. As the U.K. report says "To evaluate any use of IT, we need not only to consider its impact on specific measures of learning but also to assess the significance of the kinds of learning it supports in relation to the overall goals of education and teaching" (BECTA, 1998, p. 34).
Further to the way that knowledge is represented, in various forms of computer-assisted learning, there needs to be a consideration of whether the system helps the learner to discover the mappings and equivalencies across different representations. Part of a consideration of a model of knowledge representation also includes being able to see the logic for the sequential organisation of tasks, in other words the principles on which they are based. One of these is whether it is educationally more effective to ensure mastery of 'basic skills' first and then to seek to teach how these may be applied to more 'authentic' problems, or whether it is better to introduce learners first to authentic problems and later help them to master the procedures, techniques or conventions needed? This is a very important question as SuccessMaker, for example, appears to operate on the former principle and leave the latter to the teacher.

Allied to this is the issue about whether, when introducing a new topic or concept area, it is best to start with a range of concrete examples and then to move on to more abstract or general ideas, or more advisable to start with the general ideas and move on to specific examples. Further, there is the question of whether different learners profit differentially from a decision to operate according to one route rather than another. Perhaps the broad evaluative question suggested by these issues is whether the computer-assisted learning software supports more than one learning pathway. Individual learning opportunities for children involve more than levels and pace. They also include preferred learning style. The one size fits all philosophy may end up fitting no one.

Then there are questions concerning the way in which any management system that places students and advances them according to performance, derives its norms. If a system classifies learners according to some 'average' or expected level of performance, on what population of learners are such measures based? More importantly for our local context, can we expect any assessment of relative progress to be independent of the cultural and educational contexts from which any such norms are derived?

Finally, much more information is needed on the best approaches to adopt when trying to integrate ILS with the curriculum. This information needs to have a basis in both theory and empirical research. It is important to highlight the fact (and the BECTA Report does this) that experienced teachers know and hold beliefs about the proper responses to these issues but they need to recognise and value the relevance of what they know to evaluating whether or not to adopt an ILS or, for that matter, any form of CAI.
1.0 Recommendations

1. Computer-assisted learning, especially an integrated learning system, is a relatively costly option as an intervention. Given that, overall, its effectiveness falls below the mean of all other types of intervention, schools need to think carefully about cost effectiveness, notwithstanding other value added features of technology they may perceive (see Part B, Section 1.4, pp. 42-43 in this report).

2. ILS have shown best results in basic skills acquisition, particularly in maths. There is evidence of difficulty with the application of these skills so teachers need to plan work specifically to help students apply the basic skills they are learning (see Part A, Section 2.7, p. 35 in this report).

3. There is little evidence of generalised gains in reading from time on ILS. Schools should weigh up cost effectiveness against alternative interventions in the literacy area (see Part B, Section 1.3, pp. 41-42 in this report).

4. Consideration should be given as to whether the pedagogical approach of the software is compatible with classroom methods. Similarly, consideration should be given as to whether the content of the ILS curriculum matches that of the school curriculum and whether the type of knowledge and skills and the way they are assessed on the system matches curriculum assessment tasks (see Part C, Section 1.2, pp. 46-51 in this report).

5. As integration with the classroom program has been shown to be related to better outcomes, there needs to be planning to ensure that work on and off ILS is complementary and integrated (see Part C, Section 1.2, pp. 48-50 in this report).

6. As the validity of assessment from an ILS is questionable, schools need to make judgements about viability and progress of students against appropriate external measures (see Part A, Section 2.7, p. 35 in this report).

7. Different configurations of resources and allocation of personnel can affect outcomes so schools should be cognisant of these when planning implementation (see Part C, Section 1.1, pp. 45-46 in this report).
References


## Appendices

### Appendix 1

Effect Sizes for a Range of Typical School-based Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>No. of Effects</th>
<th>Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reinforcement</td>
<td>139</td>
<td>1.13</td>
</tr>
<tr>
<td>Students' prior cognitive ability</td>
<td>896</td>
<td>1.04</td>
</tr>
<tr>
<td>Instructional quality</td>
<td>22</td>
<td>1.00</td>
</tr>
<tr>
<td>Instructional quantity</td>
<td>80</td>
<td>.84</td>
</tr>
<tr>
<td>Direct instruction</td>
<td>253</td>
<td>.82</td>
</tr>
<tr>
<td>Acceleration</td>
<td>162</td>
<td>.72</td>
</tr>
<tr>
<td>Home factors</td>
<td>728</td>
<td>.67</td>
</tr>
<tr>
<td>Remediation/feedback</td>
<td>146</td>
<td>.65</td>
</tr>
<tr>
<td>Students' disposition to learn</td>
<td>93</td>
<td>.61</td>
</tr>
<tr>
<td>Class environment</td>
<td>921</td>
<td>.56</td>
</tr>
<tr>
<td>Challenge of goals</td>
<td>2703</td>
<td>.52</td>
</tr>
<tr>
<td>Bilingual programmes</td>
<td>285</td>
<td>.51</td>
</tr>
<tr>
<td>Peer tutoring</td>
<td>125</td>
<td>.50</td>
</tr>
<tr>
<td>Mastery learning</td>
<td>104</td>
<td>.50</td>
</tr>
<tr>
<td>Teacher inservice education</td>
<td>3912</td>
<td>.49</td>
</tr>
<tr>
<td>Parent involvement</td>
<td>339</td>
<td>.46</td>
</tr>
<tr>
<td>Homework</td>
<td>110</td>
<td>.43</td>
</tr>
<tr>
<td>Questioning</td>
<td>134</td>
<td>.41</td>
</tr>
<tr>
<td>Peers</td>
<td>122</td>
<td>.38</td>
</tr>
<tr>
<td>Advance organisers</td>
<td>387</td>
<td>.37</td>
</tr>
<tr>
<td>Simulation and games</td>
<td>111</td>
<td>.34</td>
</tr>
<tr>
<td>Computer-assisted instruction</td>
<td>566</td>
<td>.31</td>
</tr>
<tr>
<td>Instructional media</td>
<td>4421</td>
<td>.30</td>
</tr>
<tr>
<td>Testing</td>
<td>1817</td>
<td>.30</td>
</tr>
<tr>
<td>Aims and policy of the school</td>
<td>542</td>
<td>.24</td>
</tr>
<tr>
<td>Affective attributes of students</td>
<td>355</td>
<td>.24</td>
</tr>
<tr>
<td>Calculators</td>
<td>231</td>
<td>.24</td>
</tr>
<tr>
<td>Physical attributes of students</td>
<td>905</td>
<td>.21</td>
</tr>
<tr>
<td>Learning hierarchies</td>
<td>24</td>
<td>.19</td>
</tr>
<tr>
<td>Ability grouping</td>
<td>3385</td>
<td>.18</td>
</tr>
<tr>
<td>Programmed instruction</td>
<td>220</td>
<td>.18</td>
</tr>
<tr>
<td>Audio-visual aids</td>
<td>6060</td>
<td>.16</td>
</tr>
<tr>
<td>Individualisation</td>
<td>630</td>
<td>.14</td>
</tr>
<tr>
<td>Finances/money</td>
<td>658</td>
<td>.12</td>
</tr>
<tr>
<td>Behavioural objectives</td>
<td>111</td>
<td>.12</td>
</tr>
<tr>
<td>Team teaching</td>
<td>41</td>
<td>.06</td>
</tr>
<tr>
<td>Physical attributes of the school</td>
<td>1850</td>
<td>-.05</td>
</tr>
<tr>
<td>Mass media</td>
<td>274</td>
<td>-.12</td>
</tr>
<tr>
<td>Retention</td>
<td>861</td>
<td>-.15</td>
</tr>
</tbody>
</table>

Excerpted from Wilkinson et al. (1999).
## Appendix 2
Studies of Integrated Learning Systems in Elementary and Middle Grades

<table>
<thead>
<tr>
<th>Location, Year, Program, Source of Data, Research Design, Comparison Group</th>
<th>Approx. # of ILS-Using Students Tested; Grade Levels; Population; Hours of Use of ILS (math + language + reading) unless noted</th>
<th>Summary of Effects Either Reported by Evaluator or Found in Evaluator’s Report</th>
<th>Effect Size (E.S.) Indicated by that Summary</th>
<th>Problem with Comparison and other Comments Regarding Study</th>
<th>Alternative Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forth Worth Parochial School 1988-89 CCC Vendor Individual Change vs. Test Normed Population</td>
<td>Est. 120 math, read, both Grades 1-7 Chapter I, parochial schools Approx. 25 clock hours</td>
<td>Spring-Spring Stanfords, Mean NCE gain 15 point, math; 12 points, reading.</td>
<td>E.S. =1.60 (math) E.S. = .70 (reading)</td>
<td>Test scores reported for only 60% of students using program. Much larger effects show for grade 1 than older grades – and what was pretest for that grade? Excluded 3% negative outliers, but no positive outliers – and all of grade 1 reading was outlier (mean NCE pretest score was compared to over 35 for other grades). Mean national increase 2 NCE for Chapter I not included.</td>
<td>Incorporated mean national NCE gain and excluding grade 1: E.S. = .80 (math) E.S. = .10 (reading)</td>
</tr>
<tr>
<td>Location, Year, Program, Source of Data</td>
<td>Research Design, Comparison Group</td>
<td>Approx. # of ILS-Using Students Tested; Grade Levels; Population; Hours of Use of ILS (math + language + reading) unless noted</td>
<td>Summary of Effects Either Reported by Evaluator or Found in Evaluator’s Report</td>
<td>Effect Size (E.S.) Indicated by that Summary</td>
<td>Problem with Comparison and other Comments Regarding Study</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Omaha Westside Community Schools, 1988-89 CCC Vendor</td>
<td>Individual Change Vs. Test Normed Population</td>
<td>Est. 170 math, read, both grades 2-6 Chapter I in 8 schools Approx. 20 clock hours</td>
<td>Spring-Spring CTBS and SRA (varied by school). Mean NCE gain 14 points, math; 11 points, reading.</td>
<td>E.S. = 1.30</td>
<td>Excluded 3% negative outliers. Huge gain for small number of hours. Questionable.</td>
</tr>
<tr>
<td>Milwaukee Public Schools, 1987-88 CCC Vendor</td>
<td>Individual Change Vs. Test Normed Population</td>
<td>Est. 600 math, read, both Grades 2-9 Chapter I in 11 schools Approx. 40 clock hours</td>
<td>Fall-Spring in 3 schools; Spring-Spring in 8 schools; ITBS. Mean NCE gain 9 points. Spring-Spring schools only: E.S. = .60</td>
<td>E.S. = .80</td>
<td>Excluded 8% negative outliers in math and 3% in reading. No positive outliers excluded. Also, high attrition: test data only only 60% of students using CCC program.</td>
</tr>
<tr>
<td>Aiken County SC 1987-88 CCC Vendor</td>
<td>Individual Change Vs. Test Normed Population</td>
<td>Est. 600 math, read, both Grades 2-8 Chapter I in 8 schools Approx. 30 clock hours</td>
<td>Different tests, Different grades (state test and CTBS), probably Spring-Spring, but unstated. Mean NCE gain 7.5 points</td>
<td>E.S. = .70</td>
<td>(Limiting to only grades with CTBS pre- and posttests makes no difference)</td>
</tr>
</tbody>
</table>
Approx. 1500 students per year; Grades 3, 5, 6
6 elementary, 3 middle schools
rural to suburban
middle class
heterogeneous district
Est. average 35-40 class hours per year.

District-reported Fall-Fall CAT scores for 4 years prior to CCC (i.e., interval of 3 years) and 3 years after (4.5 and 1.5 for grade 8). GE change in district totals between 1980 and 1986 (CCC started 1983(1/85 for grades 6+)): +1.6 years, compared to statewide change of +0.9 years.

Assuming s.d. of GE in district = 1.1 (grade 3), 1.6 (grade 5), and 2.4 (grade 8) (estimated from district graphs), and using grade-specific GE data:

Gains between 1980 & 1986. Statewide gains as comparison group:
- Grade 3: E.S. ~ .10
- Grade 5: E.S. ~ .25
- Grade 8: E.S. ~ .50
(although intervention was only 1.5 years for grade 8 students and 3 years for others).

District gains relative to state were larger PRIOR to implementation:

<table>
<thead>
<tr>
<th>Years</th>
<th>Grades 3 and 5</th>
<th>District State difference:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-83</td>
<td>.65 .47 .18</td>
<td>1980-83</td>
</tr>
<tr>
<td>1983-86</td>
<td>.35 .25 .10</td>
<td>1983-86</td>
</tr>
<tr>
<td>1986-88</td>
<td>-.23 .03 -.26</td>
<td>1986-88</td>
</tr>
</tbody>
</table>

Also, younger families in district are more educated, yielding, for example:
- 1980: grade 3: .40 GE vs. state
- grade 8: -.40 GE vs. state.

So younger cohorts (with more CCC experience) should score higher.

(1) Compare closer years: Year just before CCC vs. 2nd year (1st full year) after CCC (again using statewide as comparison point):
- Grade 3: E.S. ~ .15
- Grade 5: E.S. ~ .00
- Grade 8: E.S. ~ .00

(2) Look longitudinally at grade cohorts: 3rd-to-5th-grade GE improvement for 4 classes with 2 years of CCC vs. 2 classes with 0 years CCC:
- E.S. ~ .10

Differences between successive cohorts suggest need for longitudinal study below

<table>
<thead>
<tr>
<th>Location, Program, Source of Data</th>
<th>Approx. # of ILS-Using Students Tested; Grade Levels; Population; Hours of Use of ILS (math + language +</th>
<th>Summary of Effects Either Reported by Evaluator or Found in Evaluator’ Report</th>
<th>Effect Size (E.S.) Indicated by that Summary</th>
<th>Problem with Comparison and other Comments Regarding Study</th>
<th>Alternative Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Design, Comparison Group</td>
<td>阅读) unless noted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calvert County MD, 1983-86 CCC Independent thru LEA (same data source as above) Individual Change vs. Test Normed Population</td>
<td>653 students (including 94 Chapter I students) followed for 5 years: 2 prior to CCC and 3 after. Received CCC in grades 4-6 or 5.7.</td>
<td>Averaging 2 years of fall CAT (pre-) and 3 years (post), mean NCE gain 5.3 points; Mean for Chapter I subsample: 10.3 points.</td>
<td>Assuming s.d. of NCE was 15 points (10) for Chapter I: E.S. ~ .35 (all) E.S. ~ 1.0 (Chapt. I)</td>
<td>Of 2 years pre-CCC, the first had much lower scores all around. District-wide, gain (relative to state gain) was .60 GE between the tow &quot;pre&quot; years compared to .10 between last pre- and first post- and -.10 between 1st and 2nd post years.</td>
<td>Assume pre- was non-stable and use only 1983 as pre- data. Only data available, percentiles for median student, converted to NCE's: gained 2 percentiles from 63.5 to 65.5 (= 1 NCE). Incorporating statewide 3 point NCE gain for 3rd grade (1983) cohort over next 2 years: E.S. = -.13 For Chapter I: 3 percentile point gain from 44.5 to 48 = 1.6 NCE. Including comparison group gain: E.S. = -.12</td>
</tr>
<tr>
<td>Los Angeles, CA, 1977-1980 (3.5 years)</td>
<td>CCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Random Assignment to treatments (Different ILS subjects)</td>
<td>and Comparison with Prior or Successor Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Est. 750 students in four schools, grades 1-6 (math), 4-6 (reading), 3-6 (language). Ethnically mixed, large minority populations, two schools were Title I. 
- Est. 50 class hours per year, some in math or reading/language only; others split.

| Spring-Spring CTBS and curriculum-specific test for four years. Random assignment to math, reading/language, or mixed, keeping assignment for up to 4 years. One comparison was between users in one subject and users only in other subjects, on tests of both subjects. A 2nd comparison was with grade cohorts immediately earlier or later that did not use CCC. (A 3rd comparison was with students in 2 schools that lacked CCC, but its results are not examined here.) |

| Reported E.S. included: CTBS MATH computation: (effects CCC math-using vs. reading/language-using) 1 year: .36; 3 years: .72; (vs. cohort control) 3 years: .39 CTBS total math: 1 year: .27; 3 years: .58 vs. cohort controls: .35 CTBS total reading: 1 year: .19; 3 years: .12 vs. cohort controls: .25 CTBS total language 1 year: .19; 3 years: .26 vs. cohort controls: .28 Curric.–specific tests: 1 yr: math: .80; rdg.: .40; lang: .70 |

| Basically well-designed, but: (1) some students with limited English proficiency removed from Reading and Language groups making groups non-equivalent. (2) Heavy level of support for implementation may make results less generalizable. (3) Cohort comparison may be flawed in that opposite-subject users did better than control in grade 3-5 study. (4) Biggest technical problem is hat researchers used residual s.d. rather than raw control group posttest s.d. producing upward tilt to E.S., particularly for reading and language. (5) Because total time in subject not balanced between and language-using students, test results biased in favor of computer users for that subject. |

<p>| E.S.’s adjusted downwards by using raw control group s.d. for comparability to other studies in education using E.S. measure: E.S. math computation: (1 Yr): .26 (3 yrs): .59 (cohort): .33 D.S. math total: (1 Yr): .17 (3 yrs): .27 (cohort): .27 E.S. reading: (1 Yr): .12 (3 yrs): .06 (cohort): .12 E.S. language (1 Yr): .11 (3 yrs): .14 (cohort): .19 |</p>
<table>
<thead>
<tr>
<th>Location</th>
<th>Grades/Subjects</th>
<th>Student Details</th>
<th>Treatment Details</th>
<th>Effect Size (Fall-Spring/Fall-Fall)</th>
<th>Notes</th>
<th>E.S. (Math only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lafayette Parish LA, 1980-81 CCC LEA</td>
<td>3 to 6. Math only. Est. 25 class hours.</td>
<td>Spring-Spring CTBS. Students randomly assigned with pretest-matched students at same schools. GE gain 1.01 vs. .88 for control group. Statistically significant difference.</td>
<td>E.S. = .19</td>
<td>(Only other randomized design of CCC besides L.A. study.)</td>
<td>No change: E.S. = .20</td>
<td></td>
</tr>
<tr>
<td>Portland OR, 1981-82 CCC Independent Comparison Students’ Gains</td>
<td>5-8. Math only. Est. 25 class hours in math</td>
<td>Fall-Spring LEA assessment. Compared against Chapter I in traditional pull-out in other schools. F-ratio computed from raw scores favored CCC. Students tested after summer; difference no longer statistically significant.</td>
<td>E.S. = .30 (fall-spr) E.S. = .13 (fall-fall)</td>
<td>Traditional Chapter I taught in elementary schools, CCC in middle schools; Traditional averaged about .5 years younger grade level. But groups pretested evenly, indicating CCC group pretested relatively lower for their grade.</td>
<td>No change: E.S. = .30</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Study Period</td>
<td>CCC/LEA</td>
<td>Comparison Type</td>
<td>Student Details</td>
<td>Gains Details</td>
<td>District Performance</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Rochester NY</td>
<td>1984-86</td>
<td>CCC LEA</td>
<td>Comparison</td>
<td>1200 students, year 1; 2600 students, year 2 grades 406 at 19 schools. Pull-out, but not limited to Chapter I.</td>
<td>Spring-Spring MAT. Year 1: individual students matched to non-users; year 2, only rough group match (not successful – non-users higher pretests). No significant difference either year.</td>
<td>Assuming s.d. of NCE was 9 in reading, 13 in math (from rough data): E.S. year 1: ~ -.06 E.S. year 2: math: ~ .00 reading: ~ -.14</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>1984-85</td>
<td>CCC LEA</td>
<td>Comparison</td>
<td>700 students elem., middle in 7 schools Majority (~60%) were Chapter I or other low-achievers. Est. 25 class hours.</td>
<td>Probably Spring-Spring (but not stated) CAT compared to students at same schools not using CCC. Significantly higher gains by CCC at elem., but not middle. Gained 7 NCE reading; 5 math.</td>
<td>Assuming s.d of NCE was 12: E.S. ~ .60 (reading) E.S. ~ .40 (math)</td>
</tr>
</tbody>
</table>
Appendix 3  
Table 1

Comparison of Effective School-wide Reading Program

<table>
<thead>
<tr>
<th>Name of Program</th>
<th>Effectiveness Research</th>
<th>Cost Per Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success for All</td>
<td>CE</td>
<td>$150 year one, a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$60 year two</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$46 year three</td>
</tr>
<tr>
<td>Exemplary Center for Reading Instruction (ECRI)</td>
<td>CE</td>
<td>$30 year one</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$ Cost of training for new teachers only</td>
</tr>
<tr>
<td>Direct Instruction (DISTAR)</td>
<td>CE</td>
<td>$100 year one</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$35 additional years</td>
</tr>
<tr>
<td>Open Court</td>
<td>E</td>
<td>$69 b</td>
</tr>
<tr>
<td>Carbo Reading Styles Program</td>
<td>E</td>
<td>$38-110 year one c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$38-70 additional years</td>
</tr>
<tr>
<td>Concept Oriented Reading Instruction (CORI)</td>
<td>E</td>
<td>$25</td>
</tr>
<tr>
<td>Junior Great Books</td>
<td>CE</td>
<td>$75.50 d</td>
</tr>
<tr>
<td>Lindamood Phonemic Sequencing (LiPs)</td>
<td>E</td>
<td>$50</td>
</tr>
<tr>
<td>Project READ</td>
<td>CE</td>
<td>$54 e</td>
</tr>
</tbody>
</table>

Note 1: Table excerpted from Schacter (1999).

Note 2: The scale for Effectiveness Research is as follows:
CE = Comprehensive Evidence: Large amount of research conducted with majority of findings showing reading improvement
E = Evidence: Some research with majority of findings showing reading improvement
LE = Little Evidence: Research with mixed results
NE = No evidence: No research showing reading improvement

Note 3: Unless other wise stated, cost per student is an estimated figure based on training, materials, follow up visits, etc. It is only an estimate. We suggest you contact the vendor to attain the most accurate cost per students. a Cost includes follow-up visits and other services. b Materials cost only. Does not include training costs. c Cost does not include travel expenses for consultant. d Cost based on Basic Training. Additional costs for intermediate and advanced training. e Cost does not include travel expenses for consultant.
<table>
<thead>
<tr>
<th>Name of Program</th>
<th>Effectiveness Research</th>
<th>Cost Per Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast For Word</td>
<td>E</td>
<td>$765 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Daisy Quest</td>
<td>E</td>
<td>$8</td>
</tr>
<tr>
<td>Little Planet</td>
<td>LE</td>
<td>$160</td>
</tr>
<tr>
<td>Wiggle Works</td>
<td>LE</td>
<td>$25</td>
</tr>
<tr>
<td>IBM Writing to Read</td>
<td>LE</td>
<td>$30</td>
</tr>
<tr>
<td>Breakthrough to Literacy</td>
<td>LE</td>
<td>$625 &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Waterford Early Reading Program</td>
<td>LE</td>
<td>$950 &lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Accelerated Reader</td>
<td>LE</td>
<td>$17.5</td>
</tr>
<tr>
<td>Academy Reading</td>
<td>E</td>
<td>$20 &lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Watch Me! Read</td>
<td>N/A</td>
<td>Not available</td>
</tr>
<tr>
<td>Project LISTEN</td>
<td>N/A</td>
<td>Not available</td>
</tr>
<tr>
<td>Read 180</td>
<td>N/A</td>
<td>$100 &lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>SuccessMaker</td>
<td>E/LE</td>
<td>Highly variable</td>
</tr>
<tr>
<td>Tomorrow’s Promise Reading</td>
<td>LE</td>
<td>Highly variable</td>
</tr>
</tbody>
</table>

Note 1: Table excerpted from Schacter (1999).

Note 2: The scale for Effectiveness Research is as follows:
- **CE** = Comprehensive Evidence: Large amount of research conducted with majority of findings showing reading improvement
- **E** = Evidence: Some research with majority of findings showing reading improvement
- **LE** = Little Evidence: Research with mixed results
- **NE** = No evidence: No research showing reading improvement

Note 3: Unless otherwise stated, cost per student is an estimated figure derived from a per classroom cost per student who is in a school of approximately 500 students. If the software has a site license (which is lower than individual classroom cost) and is designed for a k-2 population we estimated the cost per student based on 200 students (40% of the schools population). If the software is designed for all grade levels, the cost was estimated on 500 students. These figures are rough estimates, and vendor should be contacted for specific prices to meet your schools needs.

<sup>a</sup> Cost includes on site training seminar, on site software installation and set up, and unlimited technical support
<sup>b</sup> Cost includes teacher training, 2 years classroom support, and 2 years of take-home materials.
<sup>c</sup> Cost includes software, computer, training, and additional classroom materials.
<sup>d</sup> Cost does not include professional development.
<sup>e</sup> Cost does not include professional development.
Table 3
Comparison of Effective Reading Tutoring Programs

<table>
<thead>
<tr>
<th>Name of Program</th>
<th>Effectiveness Research</th>
<th>Cost Per Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Howard Street Tutoring Program</td>
<td>CE</td>
<td>$80</td>
</tr>
<tr>
<td>Book Buddies</td>
<td>E</td>
<td>$20 (^a)</td>
</tr>
<tr>
<td>Reading One-to-One</td>
<td>E</td>
<td>$20</td>
</tr>
<tr>
<td>Early Steps</td>
<td>E</td>
<td>$833</td>
</tr>
<tr>
<td>Help One Student to Succeed (HOSTS)</td>
<td>LE</td>
<td>Not available</td>
</tr>
<tr>
<td>AmericaReads UNC Program</td>
<td>LE</td>
<td>$1,068</td>
</tr>
<tr>
<td>The Intergenerational Reading Program</td>
<td>LE</td>
<td>Minimal</td>
</tr>
</tbody>
</table>

Note 1: Table excerpted from Schacter (1999).

Note 2: The scale for Effectiveness Research is as follows:
- CE = Comprehensive Evidence: Large amount of research conducted with majority of findings showing reading improvement
- E = Evidence: Some research with majority of findings showing reading improvement
- LE = Little Evidence: Research with mixed results
- NE = No evidence: No research showing reading improvement

Unless otherwise stated, cost per student is an estimated figure based on a school size of 500 students and includes materials, training, and co-ordination. \(^a\) Material cost only. Does not include cost for graduate student co-ordinators.
### Appendix 4

**Advantages and Disadvantages of Various Models of Staffing of Sessions Where Pupils Use an ILS**

<table>
<thead>
<tr>
<th>Staffing Model</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| Teacher co-ordinator static in ILS room, supervising most/all classes and groups | High level of individual support available  
Opportunity for RoAs and action planning with individual pupils relating to their work on ILS  
Continuity, coherence for staff and pupils  
Flexible whole-school use possible  
Effective diagnosis can be fed back to class teacher | Subject/class teacher not directly involved, so reduced involvement and curriculum integration  
Uses one member of full-time staff  
Can limit imaginative and creative use  
Reliant on one skilled member of staff |
| Teacher and class move to specialist ILS room                                  | Subject teachers very familiar with content so there are opportunities to integrate work on/off ILS  
Can stimulate flexible usage as part of a range of learning and teaching styles  
Teachers begin to use reports information diagnostically | Timetabling can be a problem, unless used on a regular, predetermined basis (which can be inappropriate)  
Some disruption to classes moving to and from the ILS lab |
| Teacher working in own room with his/her own pupils                            | No timetabling problems – subject specialist always available  
Teacher familiar with content and reports  
Flexibility of use to support variety of teaching and learning styles: can act as a catalyst for other development  
Departmental ownership can increase teacher motivation  
May develop departmental team work in secondary school | It can be difficult to involve all members of the department  
Limited access for cross-curricular use  
Problems caused by lack of whole-school ownership – whole-school strategy needed  
May constrain the development of other initiatives |
| Part or whole of class move to ILS room with supervision by ancillary           | Flexible, whole-school usage possible  
Standard of support may be very good  
Subject teachers freed to have ‘quality time’ with smaller groups  
Can be economical in staffing  
Provides opportunities for continuity and coherence | Subject/class teachers not directly involved so reduced choice and integration  
Can limit imagination and creative use |
| ILS self-standing with supervision either intermittent or indirect             | Students who are self-sustaining may still gain from working on ILS in this way  
Very economical in staffing terms  
Can free class teachers to work with smaller groups whilst some are on ILS | Lack of supervision can devalue the use of ILS in the eyes of the pupils and result in limited learning gains  
Pupil access needs more flexible management |