

***Progress in Adult Literacy:
Putting the Record Straight.***

**A reply to the NIACE
*Commentary on the
NFER Research Report
“Progress in Adult Literacy”***

**Greg Brooks
School of Education
University of Sheffield
(formerly of the
National Foundation for
Educational Research)**

Progress in Adult Literacy: Putting the Record Straight. A reply to the NIACE Commentary on the NFER Research Report “Progress in Adult Literacy”

Greg Brooks, University of Sheffield (formerly of National Foundation for Educational Research), May 2001

Context

The report *Progress in Adult Literacy: Do Learners Learn?* by myself and former colleagues at NFER (Brooks *et al.*, 2001a) was published in January 2001 by the Basic Skills Agency, which had commissioned the research. NIACE, the National Organisation for Adult Learning, commissioned a commentary on that report from Professor Mary Hamilton of Lancaster University. This was published, under the title *Commentary on the NFER Research Report “Progress in Adult Literacy”* (Hamilton, 2001), on 9 May 2001 to coincide with a conference organised jointly by NIACE and NATFHE (National Association for Teachers in Further and Higher Education) and called ‘25 years of basic skills: failed opportunities, new start?’ This document is a response to the NIACE commentary.

General

In his foreword to the commentary, Peter Lavender, Deputy Director of NIACE, comments that ‘the research is probably the only substantial recent study on the progress made by learners in literacy programmes’. Where Britain is concerned, that sentence is accurate even without the ‘probably’. And the research was long overdue, the only previous such study in Britain having been conducted in 1976-79. Given the substantial interval, the study had to concentrate above all on estimating how much progress adults in basic skills provision make in reading and writing – anecdote and impression are no substitute for evidence of attainment. Beyond that, it made sense to try to relate any measured progress to student characteristics, and to some aspects of the basic skills teaching situation – but not much else, since an over-ambitious project might well have failed in its main objective. Many local programmes and providers already collect more qualitative information about student’s motivations and goals in attending adult literacy provision; this study was designed to complement those efforts because reliable data on attainment and progress were lacking.

So in outline the NFER study for the Basic Skills Agency:

- gathered data on adult literacy students’ progress in reading and writing
- gathered data on the students’ background characteristics and related that information to the progress results, and
- gathered data from tutors on their characteristics and on their teaching situation, professional development and views on teaching adult literacy, and related some of that information to the progress results.

The headline findings were that adults make undramatic but worthwhile progress, but that few factors associated with significantly greater progress were found.

The NIACE commentary has a few positive things to say about the research, but concentrates heavily on perceived flaws. These fall into three main categories: ‘omissions’ from the specification and design of the study, objections to the design actually used, and queries about the statistical methods and findings. The three sets of criticisms are dealt with in sections 2, 3 and 5 of the main part of this response, below, while misunderstandings arising from lack of clarity in the report are dealt with in section 4.

Overall, what the commentary and foreword attempt to do is undermine the design and methods of the study. If sustained, the criticisms would nullify the findings. The response given below should, I believe, lead independent observers to conclude that the design and methods were appropriate and reliable, and therefore that the findings are valid and accurate. While the study was in progress, many tutors told the fieldworkers that such a study was sorely needed, and were eager to hear their local results. Many students, some anxious beforehand, said they had been pleased to take part.

The crucial question then becomes, is the overall picture convincing? Officials at the DfEE have commented that the report ‘provides a significant advance on current knowledge about basic skills training and what works, and gives a new platform from which to develop further. It is a professional report, with full contextual information... [It] is a well-written and professional piece of work and ... will be ... useful’ (a DfEE statistician, 2000, in a message to the Basic Skills Agency). And the picture was received with wry recognition, apparently, by coordinators representing about half of the sampled providers at a feedback session in London.

Some independent testimony to the success of the project is available. As project director, I was invited to consultations at Educational Testing Services in New Jersey on the design of the forthcoming international study, the Adult Literacy and Life Skills Survey (ALL). The designers of that survey were keen to incorporate better discrimination within IALS Level 1 (= UK Entry Level), and our work in ‘creating the lower rungs of the ladder’ in our reading tests has been taken into account in designing a Level 1 Extension Study for ALL.

Reluctance to accept the findings of the NFER study may or may not lie behind the criticisms of the methodology, but it would be tragic if any such reluctance deflected the profession from a search for more effective ways to empower fellow citizens.

Summary

These points correspond to those in the summary on p.3 of the commentary.

- The statistical equivalence of the students who returned for the second test to the full original sample was clearly demonstrated. No practicable method of reaching the non-returners was available, especially if they had in fact dropped out of programmes.
- Since the study covered over 250 groups of students all over England and Wales the tests could not have been made separately appropriate for each group. The tests used were designed to allow students in all areas to demonstrate progress.
- No particular amount of change was expected between the two occasions of testing. The study was designed to establish how much progress was made, on average, and analysis against different lengths of time between the two occasions of testing revealed an important finding.
- The tests were chosen with care to be appropriate, and experts in the field and providers were consulted during the development process. Though there was a ceiling effect, this was because some of the students were already above Level 1, and the test was designed to measure attainment at Entry Level and Level 1. There was no floor effect.
- Many of the students tested were indeed above Entry Level, and the report actually mentions that this means that the proportion of adults with the greatest need being reached is even lower than previously estimated.
- The writing test was designed to be accessible to students in all areas. No tutor or student objected to it during testing, whether as 'trivial' or on other grounds.
- The information collected on tutors' professional development, though based on a small and not necessarily representative sample, appears to be the first attempt to gather such information on more than a handful of people. However, requests for improved professional development in this field have been frequently voiced over many years.
- The claim that regular attendance is related to progress in reading is upheld by the data, specifically for those students who attended more than 50 hours of tuition between the two tests. The claim that intensive provision is related to progress was tentative and based on inference from the data.
- Since this was a large-scale study it would not have been possible to relate progress to students' particular learning experiences. Some factors affecting progress were identified.

Progress in Adult Literacy: Putting the Record Straight. A reply to the NIACE Commentary on the NFER Research Report “Progress in Adult Literacy”

Greg Brooks, University of Sheffield (formerly of National Foundation for Educational Research), May 2001

[Referencing: All unattributed page-references are to the NIACE commentary. References to the NFER project report are given as ‘Brooks *et al.*’.]

1. Aims

We need to be clear about what the NFER project was asked, and intended, to achieve, and therefore about what was not asked or intended. Its aims were as stated in the report (Brooks *et al.*, p.8), namely

‘to investigate:

- the progress made in literacy by adults in dedicated basic skills provision offered by Colleges of Further Education ... and Local Education Authorities ... in England and Wales, and
- factors associated with improvement in literacy by adults in such provision.’

This remit was tightly focused, as it needed to be, given its scale. There are over 1000 providers of adult literacy tuition in England and Wales, and our achieved sample of 71 providers and over 1,200 re-tested students was both large in absolute terms and large enough to be representative of the major sectors. We were very conscious that the only previous comparable study in England and Wales, in 1976-79, and all three attempted national surveys in the USA, had had severe data-gathering problems. In particular, the most recent US study, the National Evaluation of Adult Education Programs of 1990-94, achieved a sample that was just three per cent of the target. In contrast, we aimed for a pre-test sample of somewhat over 2,000 and a post-test sample that would be somewhat better than 50 per cent of that, and we achieved both.

I was also very aware that projects which take on too many objectives risk achieving none of them. The project’s scope, outlined on p.8 of Brooks *et al.*, was designed to cover the main sectors of provision, and the exclusions from it mentioned there were intended to avoid complicating the data-gathering. About 80 per cent of adult basic skills students in England and Wales are in FE College or LEA provision, and just over half of adult basic skills students in England and Wales attend dedicated provision. Therefore the sectors covered by the project’s remit accounted for just under half of all adult basic skills students in England and Wales. To have tried to cover other sectors would have ramified the design for less return – and gathering data from the chosen sectors was difficult enough.

The study also had to attempt something that had not been done before, namely providing statistical results in three different forms: norm-referenced, criterion-referenced, and some form of comparison with school-level attainment. In developing instruments and statistical methods that would achieve this, we consulted a range of experts from Britain, plus Tom Sticht from California, who has 40 years' experience in the field; we also consulted providers at three briefings organised by the Agency. The tests were substantially modified during development as a result.

2. Maintaining focus

Many of the points made in the commentary say in effect that the project would have been better if such-and-such a feature had been included, and a sketch of an ideal overall design for such a project is given on pp.9-10. In every case, these extra features were not part of the remit. Moreover, they all suffer from the problem already mentioned, namely that to have made the design more complicated would have risked not gaining any useful information, and I invite anyone who feels that a survey incorporating all or even a few of the 'omitted' aspects could be conducted to attempt it.

What follows in this section is a list of the 'omitted' features. Some are just listed; in other cases I have added a comment:

- '...relat[ing] progress to student motivation or explor[ing] the wider impact on their lives' (p.4). On p.9 this is expanded to include 'learners' ... progression in terms of employment or further education, [and] ... the relationship of measured progress with the aims and motivations of either tutors or learners.' Because the project's specific and intended purpose was to assess the progress that was being made in adult literacy provision, we concentrated on measuring adults' progress in the specific skills base of literacy;
- including students who leave programmes in the investigation of progress (p.5). The comment that students who dropped out might be expected to have made less progress than those who stayed the course is one hypothesis. Another might be that they stopped attending classes because they actually made the most progress, for a variety of reasons (and we in fact speculated that this was the case for some EAL students, Brooks *et al.*, p.94). Given the fact that the results report little progress generally, the comment about drop-out is confusing. It is not as though the report shows a fast rate of progress which might have been seen as vastly over-rated. Follow-up of these people was not possible for practical reasons;
- aiming for an average interval between pre- and post-test (p.6). Since it took six months to complete the pre-testing, and post-testing had to begin before pre-testing was complete, this was impracticable. And in fact it is better that the interval varied, since an interesting result about regular attendance came out of this;
- balancing qualitative and quantitative methods (p.9). The main focus of the study was quantitative because, as has already been emphasised, progress data were lacking, and this was the new element in the research. We did collect some qualitative information,

and this valuable for contextualising the quantitative findings. However, the resources and logistics of the project would not have accommodated extensive qualitative data-gathering;

- gathering data on what staff or learners are trying to achieve (p.9). A little information on what tutors were trying to achieve was gathered through question 7.4 in the tutor questionnaire (Brooks *et al.*, pp.33, 112);
- understanding progress (p.9). The findings may be disappointing in terms of understanding why worthwhile but undramatic progress was made, but they do paint the picture clearly, and this is what the research principally set out to do. It is imperative to establish what is happening in the first instance, before reasons behind the findings can be investigated.

Desirable though all these features are in the abstract, in practice they would have made the project design several times more complicated than it already was, and that would have risked jeopardising the main objective.

3. Appropriateness of the design

The comments ‘No serious attempt was made to assess students’ writing’ (p.5) and ‘Writing was not properly tested’ (p.7) are misplaced. Each writing sample that was returned was analysed against a marking grid that gave the opportunity of charting progress over time. The prompt was straightforward and had the advantage of requiring a very similar form of writing at pre- and post-test. Since our fieldwork was on a large scale and widespread we had to use a prompt for writing which was simple, likely to evoke responses, and neutral between situations (and therefore related directly to no student’s individual situation). The scripts were analysed in great detail on the aspects which seemed to us likely to show improvement if it was occurring. The system of analysis was designed by Tom Gorman, formerly of NFER. He has over 20 years’ experience in this field, beginning with the APU Language Monitoring Project. His analytic system, though designed originally for schoolchildren’s writing, has been refined over several years for the analysis of adults’ writing through its use on parents’ writing in the evaluations of the Agency’s family literacy programmes.

Writing samples scribed by tutors were excluded from the research, which makes the comment ‘Some [tutors] wrote the response on the student’s behalf’ (p.7) invalid if meant to imply that scribed responses were analysed. In terms of writing not being scored for content (p.7), this criticism confuses the process of writing with the thought behind writing. The process of writing is accounted for in terms of handwriting, spelling, orthographic conventions, grammar and style. Whether or not people were able to have clear thoughts regarding their learning in order to write them down is not the issue here. The study was looking at whether students’ process of writing had progressed, not their process of thought.

‘The test constructed for this study bears no necessary relationship to what students had been doing in the programmes’ (p.6). It is not the case that an assessment tool used for measuring progress has to be specifically based around the content and purpose of a course. If the

purpose of a research project is to measure progress in literacy, then that is what the tests have to measure. There is also an argument for distancing the assessment from the lesson content in order to avoid teaching to the test, and giving skewed results for those classes where this has happened.

The idea that adult learners' progress could be compared with 'the number of months of literacy tuition that would, on average, result in a child moving up one key stage' (p.6) is problematic. There is confusion in the terminology here, and 'National Curriculum (NC) level' should be substituted for 'key stage'. It typically takes two years for a primary school child to progress from NC level 2 to NC level 3, and another two from NC level 3 to NC level 4. The comparison of whether adults' progress in literacy matches what would have been expected of children may be inappropriate due to the reasons behind these adults' basic skills difficulties, and the fact that children are in school full-time for much of the year.

The objection to using items originally designed for 9-year-olds (p.7) is misplaced. Items designed for children are designed to investigate children's grasp of the mechanics of reading. The only risk here is that the stimulus texts upon which questions are based may be inappropriate in terms of pitch, or points of reference. The results on these items (Brooks *et al.*, pp.48-9) clearly show that they were pitched at an appropriate level – and this fact is in itself an interesting and important finding. The point of reference problem is non-existent as any specific socio-cultural references are always excluded from texts used in international surveys of this type.

The objection to changing the basis of criterion-referencing from the BSA Communication Standards to the new QCA Standards during the project (p.7) is also misplaced. We found it very difficult to relate the items to the BSA Standards, and much easier to use the IALS-related aspect of the QCA Standards, and I can see no valid objection to this.

I have no idea what having viewed some of our findings through a 'school-focused lens' (p.8) might mean, or why our interpretations appear to be informed by experience of school literacy (p.9). In my own case, I previously had several years' experience in evaluating family literacy, including the progress made by parent; and simultaneously with the third year of the *Progress in Adult Literacy* study was directing a substantial review of research on adult basic skills for the DfEE (Brooks *et al.*, 2001b).

'The definition of progress comes from outside the programmes, through the test. It works with a notion of a "typical" programme and a national adult norm both of which are highly fictional' (p.8). If the "typical" programme is dedicated, mainstream provision, it is not fictional, and is what we were asked to investigate; if not, the meaning of this phrase is obscure to me. If the national adult norm refers to the national distribution of literacy attainment, it is not fictional either - it is derived from IALS.

'Test items from the IALS and elsewhere were adopted uncritically' (p.9). They were not - we were clear that they represented a particular view of functional literacy. However, the IALS items were the only ones from which standardised scores could be derived, and this

was the only way we could meet the requirement to estimate how good progress was in national terms.

‘No information is presented about what might be important factors for progress in terms of student characteristics’ (p.9). Some student characteristics were related to progress, and I consider those covered important, even though the commentary (p.9) describes them as ‘basic’.

‘Important information is revealed about the extent of EAL students in general provision (15%) and the range of languages spoken by students (42) but this is not picked up in the recommendations’ (p.9). This is because such recommendations would have required much more detailed knowledge of these students than was or could have been collected. The study deliberately did not include provision specifically targeted at EAL students, and it was a surprise to us to find so many in general literacy provision. However, there is clearly a case for a parallel study of EAL provision and of those EAL students who choose to attend general provision.

4. Clarifications

A number of misunderstandings have arisen because of lack of clarity in the report, and in this section I attempt to clear these up.

The seven people interviewed (p.3) were not tutors but coordinators, and their areas were picked not by judgment but from students’ results.

Qualified teacher status (pp.4, 8): though this description was introduced after most of the tutors trained, it was used in the report as a convenient superordinate term for possession of a teacher’s certificate, or a Bachelor’s degree including Education, or a PGCE. Seventy-five per cent of tutors who returned the questionnaire had qualified teacher status and over 80 per cent had a qualification in teaching basic skills (mainly City and Guilds certificates). Thus the great majority of tutors had both Qualified Teacher Status and a basic skills certificate (Brooks *et al.*, p.19 gives a full outline of Tutors’ formal educational qualifications). Objections to use of the term Qualified Teacher Status seem irrelevant.

‘Not enough information was given about the intensive sample, ... and were they all in Wales as the title of the published evaluation report implies?’ (p.5). The words ‘in Wales’ which occur in the title of that report as given in Brooks *et al.* (p.83) are a misprint. In fact, all nine intensive courses were in England, and the report on them gives the details that Brooks *et al.* do not.

‘The sampling basis for the tutor survey is not specified’ (p.5). This could and should have been stated. Briefly, each of the 263 tutors whose groups were visited for testing was sent a questionnaire. Thus the survey was not primarily intended to be representative of all adult literacy tutors in England and Wales - that would have required both a different sampling method and a much larger sample - but to capture the views of the tutors whose students had been tested, so that the students’ progress could be related to the questionnaire results. In the event, though the sample was not very large, a few important relationships between tutor

questionnaire variables and the progress data were discovered. The over-representation of full-time tutors (p.5) is almost certainly due, I believe, to the logistics of the fieldwork - full-timers were more likely to have groups running at times when the fieldworkers could visit.

5. Statistical methods and inferences

A statistician at the Lancaster University Centre for Applied Statistics was asked to give comments on relevant aspects of the report. The commentary states (p.6) that s/he 'confirm[ed] the ... concerns' of the commentary's author, and that his/her 'overall conclusion was that the statistical analysis presented is variable in its sophistication and that not enough information is presented in Appendix E to enable proper evaluation of the item response analysis presented there.'

The project statistician, Dougal Hutchison, is the Chief Statistician at NFER. He is highly expert, and has over 30 years' experience and a PhD in the field. The sophistication of Appendix E in Brooks *et al.* should have shown that the statistician involved was no novice, and the fact that less complex statistical procedures are reported elsewhere in the report only shows that he selected appropriate methods for different purposes. Dougal also spent many weeks building on the program used in IALS as a basis for our own analyses and comparisons. If the authors of the critique had requested more detail on the process he would gladly have supplied it.

None of the statistical concerns expressed in the commentary is justified.

There were two samples of students in the study, a small one of students on intensive courses which ran in March 1999, and a much larger one of students in standard provision running throughout the academic year 1998/99. Results are presented in Brooks *et al.* for the two samples combined and for the intensive courses alone, but not for the large sample. The commentary (p.5) makes an issue of the non-reporting of separate results for the large sample. The results are presented in this way because that is what was wanted. Given that the intensive students were only 16 per cent of the total, separate results for 'non-intensive' students would not have differed markedly from those for the combined samples.

Brooks *et al.* draw attention to the 'turbulence' in the students' results between pre- and post-test, using that term as a non-technical variant on 'variability': many students' scores went down, rather more stayed about the same, rather more still went up. Though the amount of variability was quite large, it was by no means exceptional, and well within the normal range found in pre/post-test designs. The commentary (p.7) says, 'A very plausible alternative interpretation is that the test itself is faulty and unreliable.' Not so: the test items had all been found to be reliable in previous studies (IALS, the 1991 IEA study of 9-year-olds, the *Lost Opportunities* study of linguistic minority adults), and we consider that they behaved with sufficient reliability in the technical sense in this study, and that the variability or turbulence in the results was therefore genuine, and not due to unreliability.

Floor and ceiling effects: There was *no* floor effect (p.5) - very few students had zero or near-zero scores. It was true that there were not enough very simple items for students at the lower end of the range (items which students with Entry level scores could consistently get

right), but this does *not* mean (p.7) that the test could not discriminate adequately at that end of the scale - it did, because there were plenty of items which students with Entry level scores had a 50 to 80 per cent probability of getting right.

There was a ceiling effect (p.5), more clearly visible in the raw scores (which have not been published) – quite a few students achieved perfect or near-perfect scores. This was because some of the students were already above Level 1, and the test was designed to measure attainment at Entry Level and Level 1. This probably limited the average progress made by intensive students, as we pointed out. However, the ceiling effect does not (p.7) explain away the finding that, in areas where no tutors had classroom assistants, students made on average no progress. Two other subgroups had initial mean scores higher than students in areas where no tutors had classroom assistants – see intensive students (Brooks *et al.*, p.52) and those looking after home/family (Brooks *et al.*, p.68).

The intensive courses ran in March 1999. They provided 60 hours of tuition in three weeks, so were indeed highly intensive. Their focus was general improvement in literacy, rather than any particular aspect. The fact that the intensive students' gain may have been limited by a ceiling effect means that no proper inference can be drawn (p.4) from the fact that their gain was less than that of the full sample. Students taking part in the intensive courses were at a higher level to begin with, so it is likely that their progress would be less easy to measure in the format used. It is therefore wrong to suggest that we made a false inference about the effectiveness of intensive courses – this was an inference (admittedly indirect) from other information in the report, especially the greater gains made by students who attended regularly.

Our statements about the relationship between hours of tuition and progress are misrepresented. We were very careful to focus only on the 51-60 hours group, and the fact that they must have attended very regularly between pre- and post-test to achieve this number of hours' attendance. We made no statement about other hours of tuition groups, because individuals in the 21-30 hours group (for example) might have attended regularly over a few weeks or intermittently over several months. We therefore did not claim, and we did not mean to imply, that 'there is some regular pattern relating progress and ... attendance' (p.7).

6. Conclusions

No study can satisfy all demands, and no test can do more than take a limited sample of attainments. Some of the criticisms are wrong, especially on the statistics. Many others are either irrelevant, or demands for unreachable perfection. Within the requirements and constraints of the study, I believe that a reasonable and sound job was done, and I stand by the findings. I therefore believe that the picture of modest but useful progress, and of a profession labouring under difficulties, is accurate and should be accepted.

References

BROOKS, G., DAVIES, R., DUCKETT, L., HUTCHISON, D., KENDALL, S. and WILKIN, A. (2001a). *Progress in Adult Literacy: Do Learners Learn?* London: Basic Skills Agency.

BROOKS, G., GILES, K., HARMAN, J., KENDALL, S., REES, F. and WHITTAKER, S. (2001b). *Assembling the Fragments: a Review of Research on Adult Basic Skills*. London: Department for Education and Employment Research Report no.220.

HAMILTON, M. (2001). *Commentary on the NFER Research Report "Progress in Adult Literacy"*. Leicester: NIACE.