



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

**DESIGNING AND USING ADAPTIVE
TESTS FOR LARGE SCALE
FORMATIVE ASSESSMENT
1999 TO 2008**

John Winkley, Bob Rainbow, Husam Baki

Designing and Using Adaptive Tests for Large Scale Formative Assessment: 1999 to 2008

John Winkley, AlphaPlus Consultancy Ltd,
Telephone: 0845 094 8825
Web: www.alphaplusconsultancy.co.uk
Email: info@alphaplusconsultancy.co.uk

Introduction

This presentation is concerned with the development and trialling of sets of adaptive computer-based and (non-adaptive) paper-based tools used primarily to assess the literacy, language and numeracy skills of adult learners in both academic and work settings in the UK.

In 2002, the *Skills for Life* Unit of the Department for Education and Skills (DfES) commissioned AlphaPlus Consultancy Ltd and BTL Group Ltd to develop suites of assessment tools to be used to support learners and teachers embarked on its *Skills for Life* 'Learning Journey'. The DfES was particularly concerned to have in place a collection of reliable and valid paper-based and computer-based 'tools' for assessing learners during the early stages of their journey towards improved literacy, language and numeracy skills.

After an intensive period of development and trialling, a collection of fifteen computer-based assessment tools and twenty-five paper-based assessment tools and associated guidance materials was delivered to the DfES, *Skills for Life* Unit in 2006. These tools have been used very extensively in hundreds of centres with thousands of learners for diagnostic and formative assessment purposes and are one of several outputs from an ongoing development of computer-based adaptive formative assessment tools. As this report will demonstrate, trials confirmed that the tools are both powerful and engaging for teachers and learners alike.

Although the expectation was that most centres would use the computer versions of the tools, paper-based versions were also commissioned for use in situations where computers were unavailable or assessors considered them inappropriate.

The Learning Journey

The United Kingdom's strategy for improving adult literacy and numeracy skills is based on a programme called "The Learning Journey". The diagram below summarises the process from an initial skills check through to the creation of individualised learning programmes and summative assessments to check progress.

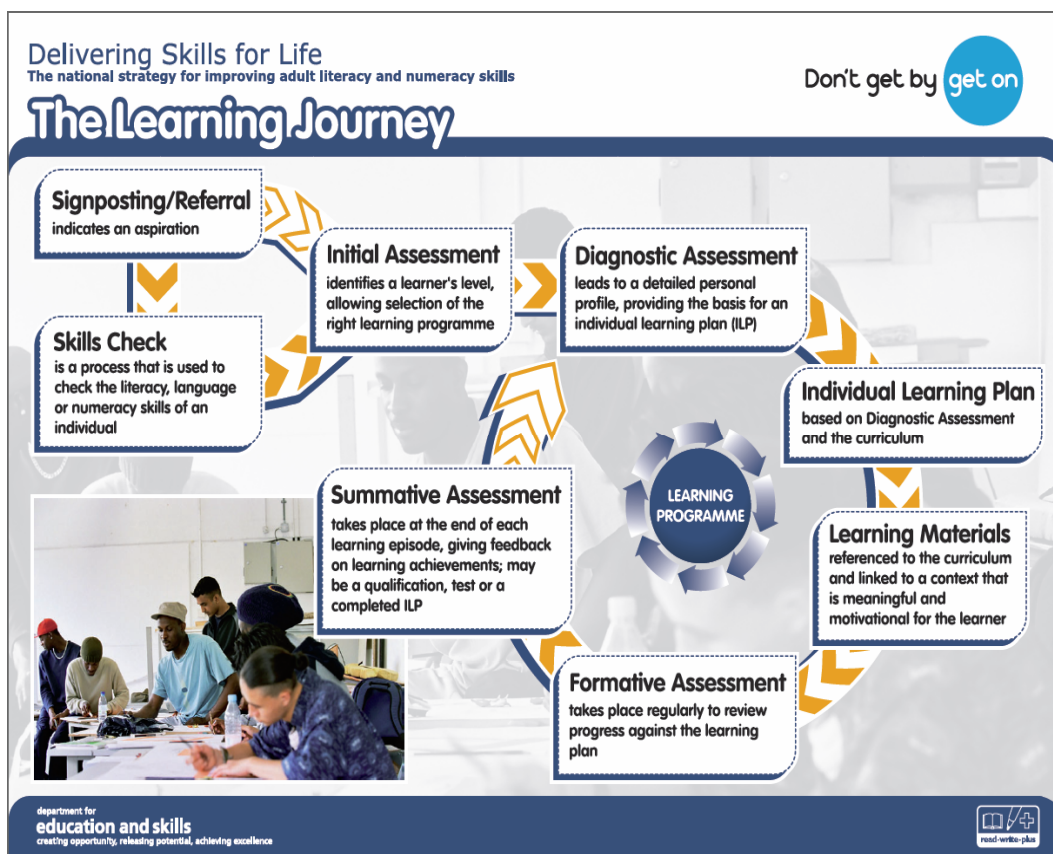


Figure 1: The Learning Journey

The full collection of materials developed to support teachers and adult learners on the Learning Journey consists of:

- Numeracy and Literacy **Skills Check** tools for use in:
 - general academic and vocational settings
 - general workplace settings
 - four sector-specific workplace settings.
- Numeracy, Literacy and English as a English for speakers of other languages (ESOL) **Initial Assessment (IA)** tools for use in:
 - general academic and vocational settings
 - general workplace settings.
- Additional guidance materials.

The Skills Check (formerly known as 'screening') tools were developed to assist 'front-line' staff (for example, in Job Centres) in assessing whether or not an adult has a literacy and/or numeracy need and would, therefore, benefit from a more in-depth assessment. The primary function of the skills check tools is to determine *need*. In particular, the DfES wanted to be able to identify whether an adult has literacy and/or numeracy skills at, or above, level 2 of the National Standards (a

critical level of competence above which Skills for Life provision is not required). The Project team's brief was to produce user-friendly tools that would give reliable results with as little testing time as possible.

The purpose of an **initial assessment** (IA) is to determine a learner's literacy, language and/or numeracy skills against a level or levels within the national standards. This information is then used to place her/him on an appropriate learning programme, at the appropriate level. The Project devised a number of initial assessment tools for both academic (e.g. Further Education colleges and Training Centres) and workplace settings, all of which are designed to be completed in approximately 20 minutes.

The adaptable, computer-based tools

As mentioned above, the project team developed a number of conventional linear, paper-based tools. These require users to attempt all questions even though they may be well beyond their capabilities. This can, of course, be distressing for many learners at the start of a course particularly those who have previous experience of failure in education. However, the computer-based skills check and IA tools are adaptive and are able to route learners automatically to questions at an appropriate level or, indeed, terminate the assessment when necessary.

Both the skills check and IA tools use adaptive algorithms to structure the tools. A simplified version of the algorithm for the IA tool is given below.

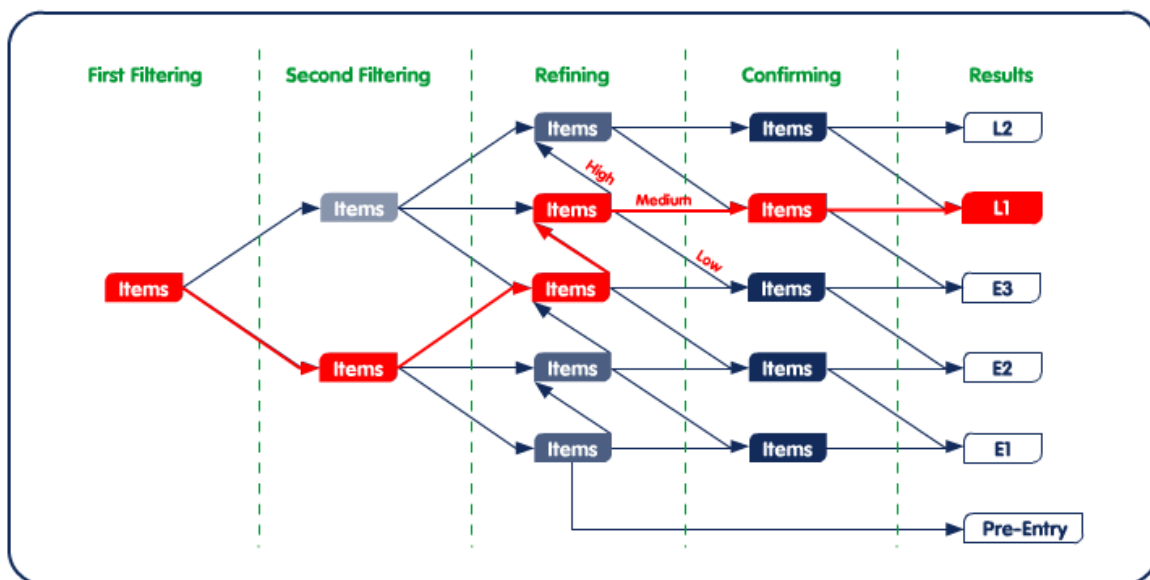


Figure 2: The Initial Assessment Adaptive Algorithm

The learner begins the assessment process on the left and attempts a given number of fairly low level questions (items) as part of the “First Filtering”. Thereafter, a “snakes and ladders” structure is used to route learners upwards, downwards or straight on according to their performance at the end of each cluster of questions (items). In the “First Filtering”, questions begin at Entry 1 and end at Level 1. Those

candidates who successfully achieve above the threshold mark for this section are routed upwards to attempt Level 1 and Level 2 questions.

Those who are unsuccessful are routed to Entry level questions in the “Second Filtering”, and so on. In this way it is possible for learners to climb to the highest level, even after a poor start. The level achieved in the “Confirming” section is the final, approximate level. This is instantly displayed for learners and assessors and indicates questions correctly and incorrectly answered. This information can be stored and/or printed off.

The adaptive nature of the tools makes them less threatening for learners than conventional tests as they are always asked questions broadly within their own skill level. All of the questions in both types of tool are aligned to assessment criteria from the Adult Core Literacy and Numeracy Curricula.

Innovative question types

A major goal for the project was to provide assessments that were clear, quick and easy to use, innovative in design and stimulating and user-friendly for learners. At the same time, the questions had, of course, to be valid and reliable.

The evaluation reports for the project indicate that e-assessments are popular with all learners but particularly so with younger adults. They welcomed the use of sound, colour and onscreen images. A number of innovative assessment techniques were used in the tools which not only improve engagement but also allow additional skills to be assessed. For example, sound was used to assess listening skills in the literacy tools and was also used in the numeracy assessments to read out questions, thus overcoming the perennial problem for those learners who “could do the sums if they understood the question.”

In addition to conventional multiple-choice questions, a wide variety of assessment types, all of which are computer-marked, have been used in the tools:

- highlighting text or other information
- dragging-and-dropping words/images/numbers
- gap-filling
- proof-reading and correcting text
- manipulating graphs and charts
- measuring (using onscreen rulers, thermometers, measuring cylinders)
- weighing
- animation (e.g. tallying and graphing traffic flow).

Through careful matching of the technical capabilities to the assessment requirements, the breadth of coverage of assessment is greatly increased. Two examples are included below:

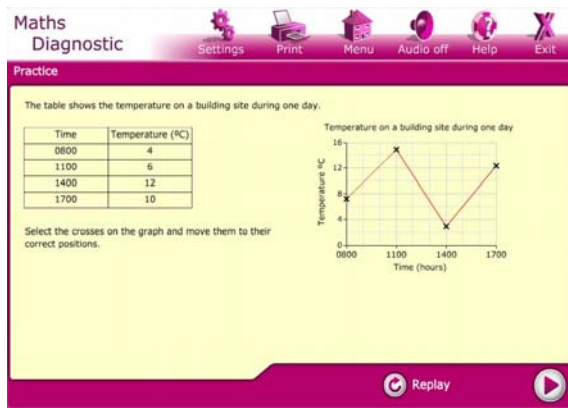


Figure 3: Manipulating graphs and charts

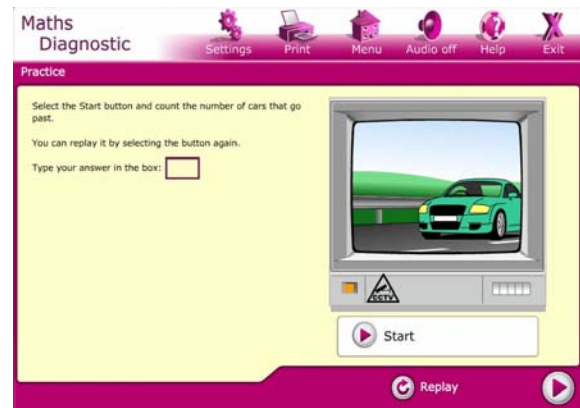


Figure 4: Animation – tallying traffic flow

These techniques also proved popular with learners who enjoyed the variety and challenges they offered. They also preferred them to writing on paper!

The advantages of e-assessment

Computer-based tools offer a number of advantages over equivalent paper-based instruments for skills checking and initial assessments. e-Tools can:

- be adaptive to learner performance, maintaining the level of challenge at the right level
- be marked automatically by the computer and provide immediate results
- produce a detailed profile of a learner's performance available for immediate review which may be integrated into electronic learning planning activities
- be less threatening for learners (by being adaptable and by avoiding the connotations of "exams" introduced by using black and white test papers)
- provide 100% accurate marking, every time
- use rich authentic media content – newsprint, images, maps, forms, articles, and colour, sound, moving images, animation and video.
- include onscreen embedded tools such as a calculator
- provide support for people with certain disabilities (large size print at the touch of a button for visually impaired candidates; use of audio; use of colour button for dyslexics; use simple mouse or key operations for physically disabled learners)
- (i.e. once the prototype has been developed and the "engine" to drive the programme has been produced)
- generate test data easily and efficiently (for trials and ongoing performance monitoring).

Both the computer-based skills check and initial assessment packages developed for this project contained many of the features listed above. A number of them are worthy of more detailed comment.

Contextualised tools

The Project developed sets of “Standard” tools, designed for general learning environments such as Further Education colleges and community learning centres. These tools are populated with questions that relate to general, everyday-life and activities. The Workplace tools that also have been developed assess broadly the same criteria as those in the Standard tools but the questions are set in general workplace contexts. In the suite of sector-specific tools, the questions have been adapted to the relevant workplaces. A separate ESOL skills check tool is also available. The ESOL tools use more multi-cultural stimuli and place a greater emphasis on speaking and listening skills. The computer versions of the workplace and sector tools incorporate images and graphics taken from the workplace and, therefore, stimulate learners, who can relate what they see onscreen to their own working environment, and thus see the relevance of the questions.

Trialling the tools

Prior to trialling the tools with learners, the tools underwent rigorous beta testing¹. Testing the functionality of tools with an effectively infinite variety of learner paths and scores (the algorithms are more complex than the example shown above) presents technical challenges for test engineers, as well as the expert review of fitness for purpose which precedes user trials.

The tools were trialled in approximately 300 centres and with several thousand learners². Traditional protocols for trialling tests were not applicable – in tests where grades are assigned to mark bands (eg Grade A 68-100%, Grade B 57-67%, etc) the processes revolve around comparison with previous calibrated benchmark tests, and expert inspection. For minimum competence (pass fail) tests, the most commonly used technique in the UK (Angoff, Modified Angoff) relies on combining judgements from a panel of experts about the level of difficulty of each question. Candidates on adaptive assessments follow a variety of paths and hence the “output point” of the algorithm determines the level (the assessments developed here provide one of 6 levels). Trialling was therefore based on comparing output level from the test with other measures of level (teacher review, other tests) and basic forms of item level analysis. Correlation with predicted results was very high (for 90% of cases, less than 1 level difference), resulting in the assessments being published and passing into common use.

Further work – extending the performance of adaptive assessment

The initial assessment and screening tools are designed to provide indications of level only. Since the project ended, we have been working on using the same algorithmic techniques to produce not only an aggregate level, but also detailed feedback about the learner’s level in the individual skill areas which make up that level. An example of such output is included in the screenshot below.

¹ 1. Data are available from John Winkley, AlphaPlus Consultancy Ltd, on request.

² Trials data is held by the DfES.

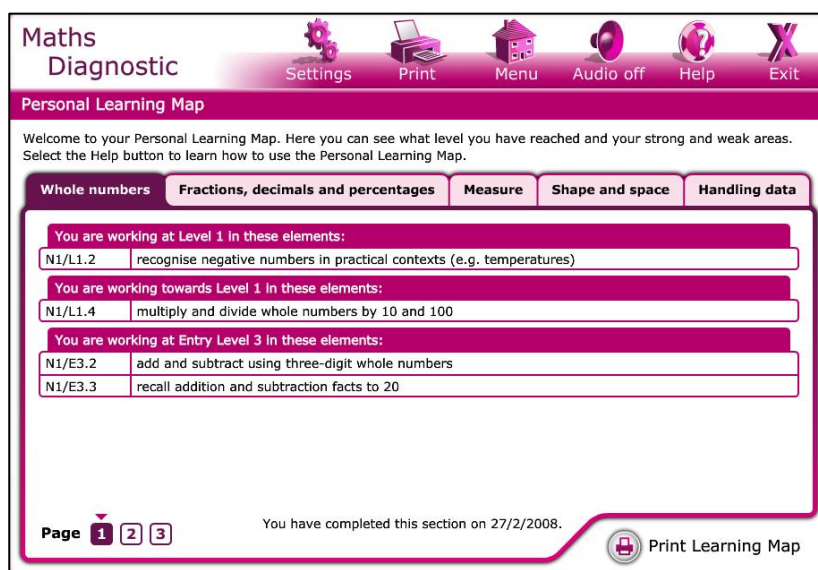


Figure 5: Personal Learning Map

These tools form the core diagnostic assessment tools (cf the Learning Journey diagram presented earlier) for Learndirect's Skills for Life assessment programmes which are used by hundreds of thousands of learners.

A successful product for this purpose faces a number of challenges. The curricula for the Skills for Life skills are very broad, comprising approximately 150 curriculum topics across the 5 levels in each subject. If we assume that:

- a) no information exists about the learner's level when starting the diagnostic
- b) at least 3 questions need to be asked in each topic to be confident of making a *competent/partially competent/not competent* decision

This suggests a maximum of 450 questions for each subject (depending on which adaptive path learners follow) and at least 250 questions for the more common paths. At progress rates (from trials) of about 1 question per minute, this implies a 4 hour diagnostic assessment for each subject which is too long, both for the learning programmes that providers offer, and for learners who are likely to want to revisit the tests, using them for formative as well as diagnostic purposes.

The new adaptive algorithms we devised reduces this workload by making judgements about competence at a lower level contributing skill, if a higher level skill which uses it is demonstrated successfully.

In Figure 6, dependency (linking) lines are shown as black lines. In this case, a user has chosen to start the diagnostic at Level 1 (Grey Box 5) and is required first to answer questions for each curriculum element at that level (Boxes 5, 6, 7, 8, 9 and 10).

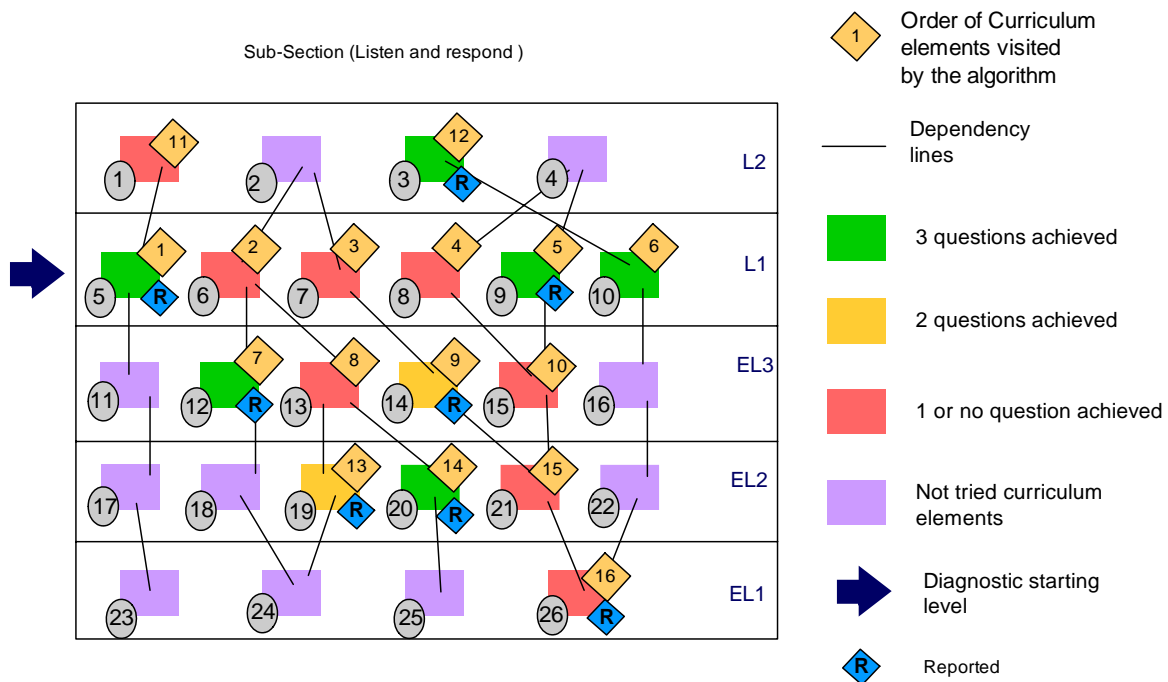


Figure 6: An Adaptive Algorithm for a Diagnostic Assessment

Each coloured box represents a topic with three questions. Learners answer all the questions in topics at their starting level, and then the algorithm presents them with selected questions at other levels based on their performance.

Getting users to start at what they or their tutor deem to be the appropriate level should reduce the time taken for the diagnostic. It is important that the starting level is determined as accurately as possible. To ensure that this is the case, as stated previously, the starting level should be determined by one or more of the following:

- an initial assessment tool
- the result of previous curriculum elements achieved at a lower level, using this diagnostic tool
- advice from a teacher/supervisor
- achievement of a similar qualification at this level.

The diagnostic algorithm works as follows:

- each curriculum element box is numbered (Grey Oval numbers 1 – 26). These boxes are populated with 3 questions
- boxes are coloured as follows:
 - green for boxes for those in which the user achieved 3 marks (e.g. Box 5)
 - yellow for boxes for those in which the user achieved 2 marks (e.g. Box 14)

- red for boxes for those in which the user achieved 0 or 1 mark (e.g. Box 13)
- Purple is used for boxes that were not attempted (e.g. Box 1) or where the algorithm would predict that the user would pass, if attempted (e.g. Box 11), and thus not relevant.
- In the example in Figure 6, the algorithm first sets questions at the user's starting level, Level 1, i.e. questions are taken from boxes 5, 6, 7, 8, 9 and 10. Depending on the user's performance, the algorithm routes the user to questions from one or more boxes from 12, 13, 14 and 15, (if they fail, score 0 or 1 on the linked Level 1 elements) or from boxes 1 to 4 at level 2, (if they achieve 3 marks on any of the linked elements at Level 1).
- This algorithm focuses the learner on challenging questions, and avoids topics where the learner has already (indirectly) demonstrated competence through a related skill at a higher level.
- On completion of the assessment, the learner is presented with a "spikey" profile, showing their performance in each topic, as illustrated in Figure 5.

What has been learnt?

The project work described has been groundbreaking in both its development and deployment of innovative e-assessment techniques and tools. The algorithms that underpin the tools and the questions within them represent a significant breakthrough in this type of testing and most adult learners have responded well to these fresh approaches to assessment and learning.

Learners have responded well to these tools. For example (in the case of Literacy), 60% of learners prefer the on-screen test after a first sitting with no preparation, 24% have no preference for on-paper or on-screen, and only 16% prefer paper. Most have found the tools easy to use, including learners with special access requirements.

The tools can be downloaded from <http://www.toolslibrary.com/> .

With special thanks to BTL Group Ltd (the technologists) www.btl.com and Learndirect www.learndirect.com