SECTION SIX

THE IMPLEMENTATION OF THE TEST EQUATING MODEL

6.1 Implementation of the test equating procedure

The overarching implementation model involves each State/Territory taking one Home Test (for example NSW taking BST) and one Foreign Test (for example NSW taking MAP). The procedures and timeframe for implementation are detailed in section 10. The implementation of the test equating model outlined in this report has the following elements:

6.1.1 Consultation

Consultation with States/Territories began in Phase One of this project. Consultation should continue throughout Phase Two. Prior consultations with relevant groups within States/Territories should be conducted regarding key stages of Phase Two.

6.1.2 Sampling

The simple random sampling equivalent sample size for the whole project is calculated on the basis of an achievable standard error of 0.03 as detailed in section 3. In implementing the plan, intact classes from States/Territories are anticipated to be involved in the sample. The detailed procedures used in the sampling of schools and classes within each State/Territory should be determined in consultation with each State/Territory at the beginning of Phase Two.

6.1.3 Test period

All tests, both Home and Foreign, should be administered during the week beginning 20 October 1997 and ending 31 October 1997. This period occurs during Term 3 for Tasmania and Term 4 for all other States/Territories. With the exception of Hobart Show Day on 23 October (Thursday) in Tasmania, this period has no overlap with any public holidays in the States/Territories. The short period of time within which all candidates would take the tests should ensure there is minimal difference due to gain in literacy as a function of time differences in testing.

6.1.4 Test order

All States/Territories should administer the Home Test before administering the Foreign Test. No two tests should be administered on the same day. This is to minimise errors introduced due to student fatigue.

6.1.5 Test form

With the exceptions of DART, where the 1994 copyright version would be used, and MAP, where the 1995 Urban version would be used, the test form should comprise the 1996 version of BST, LAP, NET and MSE. The whole test in its original format could be used as test forms. This is to ensure authenticity of tests. For example, in tests where a single stimulus material is used for several related test items, selection of a subset of the items would require detailed item analysis which would not be required.

6.1.6 Test time

Time allowed to complete the test should follow that specified for the original test. Negotiation with schools and teachers on the best class period in which to administer the test (for example, students would be given as long as they need to complete the test where the test has no time limit) would be through States/Territories and should take place at the start of Phase Two.

6.1.7 Test printing, collating and distribution

As far as feasible, the original agencies for printing, collating and distribution of test forms should be involved. Details for each State/Territory are given in section 9. Arrangements for printing, collating and distribution of tests should take place at the start of Phase Two.

6.1.8 Test administration

Individual States' and Territories' internal arrangements for test administration are detailed in section 8.

6.1.9 Scoring

As far as feasible, the original agencies for scanning and scoring of test results should be involved. Arrangements for scanning and scoring of test results should be made at the start of Phase Two.

6.1.10 Marking for extended response items

Some test items require students to write between one line of text and a paragraph, occasionally more. Marking of such extended responses can only be done one at a time by experts, and is thus resource-intensive. Such scripts should be marked by the agencies associated with the tests. Where this is not feasible, the scripts should be marked centrally.

6.2 Principles underpinning the selection of tests for States/Territories

Two principles underpin the selection of tests for States/Territories. These are (a) the authentic principle and (b) the minimum error principle.

6.2.1 The authentic principle

According to the authentic principle, tests are selected for States/Territories such that those aspects of literacy (for example, writing) deemed important by the State/Territory are maintained. This is ascertained by comparing and contrasting the test contents and pairing Home and Foreign Tests accordingly. Table 6.1 summarises the reading/writing components of the Home and Foreign Tests. It is assumed that those States/Territories where no or minimal writing (for example, NSW) is involved in their Home Test would be given a Foreign Test with no or minimal writing component (for example, NSW has MAP, the writing component of which involves teacher-selected student writing portfolios only). This is illustrated in table 6.1.

Table 6.1 Reading/Writing components of Home and Foreign Tests

Test	State		NSW	Vic.	Qld	SA	WA	ACT	NT	Tas.
	Home test		**BST**	**LAP**	**NET**	**BST**	**MSE**	**DART**	**MAP**	**DART**
	Test Version		1996	1996	1996	1996	1996	1994	1995 Urban	1996
BST	Read	-	**Read**			**Read**		Read		
LAP	Read	Write		**Read & Write**					Read & Write	Read & Write
NET	Read	Write		Read & Write	**Read & Write**					
MSE	Read	Write			Read & Write		**Read & Write**			
DART	Read	Write				Read & Write	Read & Write	**Read & Write**		**Read & Write**
MAP	Read	Write	Read & Write						**Read & Write**	

6.2.2 The minimal error principle

According to the minimal error principle, the Home Test from a small State/Territory (for example DART of ACT) is linked to a Foreign Test with a large database (for example BST with database from NSW and SA). This is to ensure small calibration error in the linking process.

6.3 Establishment of links between tests

Links between tests are developed through common-person-common-test matrix sampling design. Linkages are achieved by each candidate taking two tests, and each test being attempted by candidates from at least two States/Territories. This is illustrated in figure 6.1.

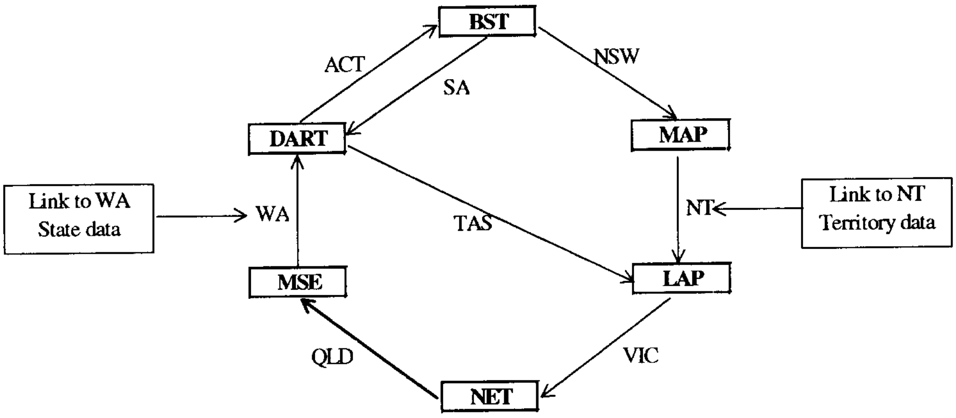

Figure 6.1 Test links

Two properties of IRM equating methods enable links, either directly or indirectly. These are (a) the symmetry property and (b) the equity property.

6.3.1 The symmetry property

The symmetry property of test equating means that once equivalences between two tests has been established, it does not matter subsequently which one is used as the base. For example, if it has been established that a score of 10 on Test X is equivalent to a score of 20 on Test Y, then a student scoring 10 on Test X is equivalent to scoring 20 on Test Y and another student scoring 20 on Test Y is equivalent to scoring 10 on Test X. The symmetry property automatically rules out regression as a test-equating method.

6.3.2 The equity property

The equity property of test equating means that if two tests are equated, then (a) it is a matter of indifference to each examinee which test was attempted and (b) equivalence is established independent of the group used. (Morris, 1982)

The mechanics of IRM equating involves three steps: design, calibration and equating.

6.3.3 Design

Determination on data collection design for equating is discussed in sections 3 and 5.

6.3.4 Calibration

Calibration of test items on the same literacy scale is discussed in section 5.

6.3.5 Equating

Equating involves using the relationship between raw score to literacy score on each of the two tests to be equated to establish the equivalences between the raw scores of the two tests. It is possible to perform steps 6.3.4 and 6.3.5 together.

6.4 Compilation of the raw data

Raw data from each State/Territory have to be centrally processed by the Home State through a Phase Two research team. For example, all BST scripts are shipped to New South Wales to be scanned and all LAP scripts shipped to Victoria to be scored and marked. This is shown in table 6.2 below. Data compilation involves scanning/marking the raw data at the home State/Territory, storing the data in electronic format, checking the scanned data/mark, and pooling data from all sources into a large data file.

Table 6.2 Data compilation responsibilities

Test	Data source	Scanning/marking responsibility
BST	NSW, NT, SA	NSW
LAP	NT, Tas., Vic.	Vic.
NET	Qld, Vic.	Qld
MSE	WA, Qld	WA
DART	ACT, SA, Tas., WA	ACER
MAP	NSW, NT	ACER

□ Compilation of the raw data is performed in two steps: (a) scanning of multiple-choice items (for example, BST), and (b) marking of writing components of tests (for example, marking the writing component of MSE).

6.4.1 Mapping of profiles onto the common metric of tests

In the analysis according to the Rasch models, all items from all tests are treated as a common item pool. It is this feature that provides the common location of all items on the continuum of achievement. The items within each identifiable strand of literacy (for example, writing and reading) can be located on separate scales, but still be referred to the common metric.

The cut-off points associated with the National Profiles, where they are used, can be estimated by locating the items on the scales and matching the items to the National Profiles. Although all States/Territories do not use the National Profiles, there is sufficient commonality among the versions of the profiles and among the test forms which are reported in terms of profiles to provide a common eight-point scale. The Victorian, Tasmanian, Australian Capital Territory, Queensland and Western Australian applications of the profiles can be integrated to establish the relevant anchor points.

This step would then enable direct comparisons of the States/Territories in terms of outcomes based on distributions of students over the National Profile levels. An example of this type of chart is presented in figure 6.2 below illustrating the percentage of studenι cohorts at each level of the National Profile for each State/Territory.

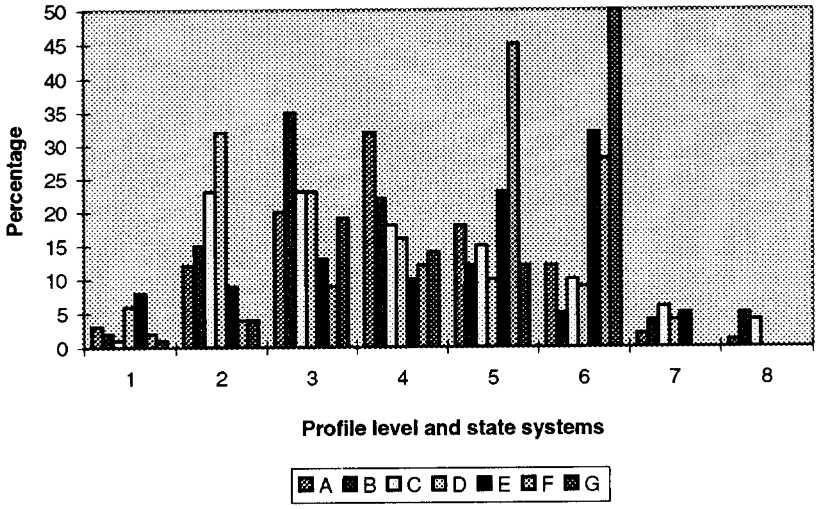

Figure 6.2 Example distribution over National Profile levels

6.5 Recommended model for future monitoring

6.5.1 The model

The model proposed in this project can be generalised across the curriculum and over time. Each State/Territory can engage in developing items trials, by agreement with another State/Territory. A central coordinating body can be set up to establish procedures that ensure that test dates are archived and used for cross-system equating.

For example, Queensland, New South Wales, Tasmania, Victoria and South Australia currently use neighbouring States/Territories in calibrating items for their

Home Tests. Of the three outstanding states, Western Australia has been the host state for calibration trials of Victorian items. The ACT has adopted the practice of using commercially available materials. Only the Northern Territory differs in this process. It appears that no items are calibrated for other systems in the Northern Territory, and the Northern Territory system develops and calibrates its own items with the Territory. If the Northern Territory system were encouraged to trial in another State/Territory, a national network of calibration trials could be established. Data from these trials could be archived and, if available, used in a variation of this project's proposed model, to link all State/Territory cohort tests without the need for any further testing in schools. This approach would preserve the integrity of State/Territory testing programs, and avoid the imposition and expense of further testing programs.

The nationwide operation outlined above goes beyond States' and Territories' normal test administration procedures and is most efficiently done by the central coordinating body suggested above.

The general procedure of the operation would include the following:

□ a central national test equivalence database is to be established from the data collected in Phase Two.

□ States/Territories would forward test results to the central coordinating body after each test session.

□ Calibration would be performed by the central coordinating body for States/Territories using the national database.