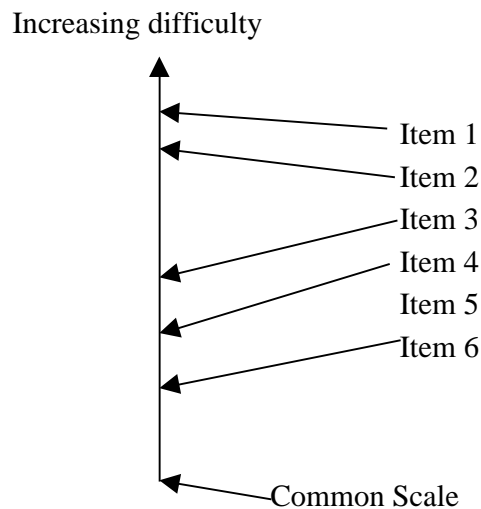


## Using Rasch Analysis to Scale TCE Subjects

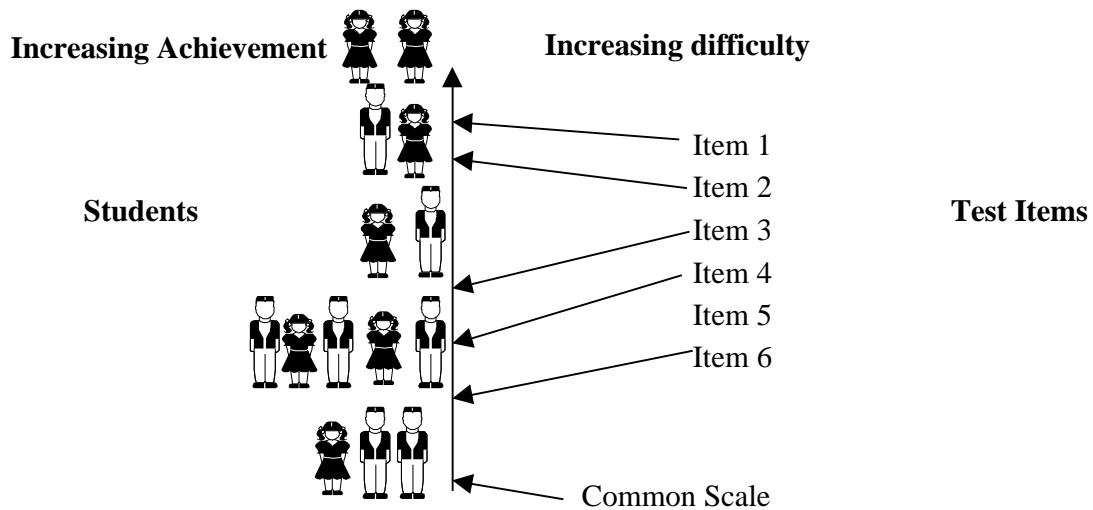
### ***Rasch Basics***

Rasch Analysis was developed to check the difficulty of different test items and calibrate them on a single, common, scale. It assumes that each item measures an underlying common characteristic or trait. By analysing how students perform on a range of test items it is possible to arrange the items on a "difficulty scale". For example, for six items of a test, the difficulty might be as shown on the common scale below:



If, in the set of tests shown in the diagram above a student gets items 3, 4, 5 and 6 correct and items 1 and 2 incorrect then it is possible to assert that the student's ability lies somewhere between the difficulty of item 2 and item 3. This represents the ideal situation where a student's achievement can be accurately located on a common scale. In the real world a student is likely to get some difficult items correct and some easier items incorrect. Rasch analysis uses the probability of a student getting a question correct or incorrect to determine both the achievement of the student and the difficulty of the items on the same scale.

Thus, at the same time as test items are calibrated on the common scale, Rasch analysis, because student performance on the items is known, is also capable of translating student achievement onto the same scale. Rasch terminology refers to this as student ability but "Rasch ability" is actually a measure of the probability that a student will be successful in correctly answering questions whose difficulty is known. Rasch therefore not only calibrates item difficulty but also ranks student achievement. This is graphically depicted as below:



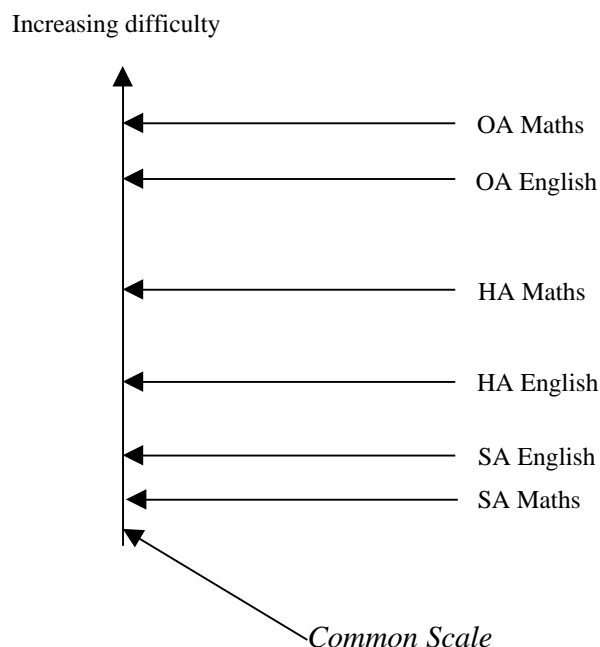
The power of the Rasch model stems from the fact that the scale is independent of the sample of students taking the items and the fact that it can be applied when the difficulty of the items being used for assessment are not equally spaced. This is demonstrably true for TCE criterion based assessment. It is this characteristic that makes the Rasch model valid in a criterion based assessment environment unlike other methods used around Australia that assumes or artificially generates a normal distribution of student achievement.

Rasch also calculates a "fit statistic" for each item so that, it is possible to establish whether a test item is measuring a common underlying characteristic. For example, a question on Shakespeare in a numeracy test should show up as an item that does not fit, as it is unlikely to be measuring the underlying trait "numeracy". The need to ensure that each item is measuring the same underlying trait is important in using Rasch to calibrate or equate test items. Initially the methodology was applied to items that are "yes/no" but it is easily modified to use on graded assessments as in the TCE.

### ***How Rasch Might be Applied to Subject Scaling***

Taking a set of different test items that measure an underlying characteristic and placing them on a common scale is exactly what we require for inter-subject scaling. The test items in this case are the subject assessments and the underlying characteristic that is being measured is "general academic ability" or "merit to enter university" of students. Once subject assessments are placed on a common scale they are directly comparable. Without first placing them on a common scale they are not comparable and therefore should not be used to aggregate scores.

Instead of considering test items, the overall assessment of a whole subject can be scaled using Rasch methodology. It is proposed to take the awards, Satisfactory Achievement (SA), High Achievement (HA) and Outstanding Achievement (OA) for each subject and equate them to scores on a common scale as shown in an example below.



Instead of using fixed thresholds for awards as is now the case (SA 1-8, HA 9-16, OA 17-20), the Rasch method takes into account all of the subjects undertaken by a student and adjusts the award threshold positions on the scale according to the difficulty of the subjects.

By knowing the scaled thresholds for the awards it is possible to interpolate (fill in) the scores in between. Hence, for each student in each subject, it is possible to calculate a score on the common scale. This ensures that scaled subject scores are directly comparable. Then, and only then, is it legitimate to add different subject scores together.

TASSAB will first use Rasch Analysis to calculate the award cut-off points for all subjects and then calculate the scaled score for each subject. This will be adjusted to give a score similar to the old 20-point score. The Tertiary entrance score will then be calculated by taking the best 5 scaled scores including at least three from Year 12.

Note that, because the same common scale is used every year, it is possible to compare student performances between Year 11 and Year 12.

TASSAB has sought expert advice on the model from the Australian Council for Education Research (ACER) and Professor George Cooney who is Chair of the NSW University Admissions Centre. ACER indicates that the method is a definite improvement on the current situation although they believe that the use of the student achievement measure for tertiary entrance admission would be the preferred option. The TASSAB Board has rejected this approach because using the "Rasch Ability" scale:

- is likely to encourage a narrow focus by reducing the number of subjects studied by students wishing for a high tertiary entrance ranking;
- is a radical change from current practice;
- is difficult to explain to stakeholders; and
- makes it difficult to report progress for Year 11 students undertaking pre-tertiary subjects.

Professor Cooney has indicated that the model is an appropriate scaling model to use with TCE examination data and that it is consistent with current test theory and the way that awards are assigned. In short he "likes the model" and believes that "it is very suitable for the Tasmanian situation".

### ***Statistical Checking of Internal Assessments***

As mentioned above, not only does the Rasch Analysis place all of the subjects on a common scale but also places all students on a common ability scale. This will provide TASSAB with an additional quality management tool by providing a baseline measure to monitor internal assessments to determine if they fit the pattern predicted by the student achievement profile of classes.