Archived at the Flinders Academic Commons

http://dspace.flinders.edu.au/dspace/

This is the publisher’s copyrighted version of this article.

The original can be found at: http://iej.cjb.net
The Development of Scales to Measure Students’ Teachers’ and Scientists’ Views on STS

Debra K Tedman
Flinders University of South Australia  dtedman@stjohns.sa.edu.au

John P Keeves
Flinders University of South Australia  john.keeves@flinders.edu.au

The starting point for this work on the development of scales was an existing instrument concerned with Views on Science, Technology and Society (STS) which had been prepared in Canada. This Australian study developed scales to measure views towards Science, Technology and Society, and, it was necessary initially to specify scores to the alternative responses or views for each of the statements included in the scales used in this study. The initial scores or codes for the scales were based upon preliminary analysis and the researcher’s judgment derived from a review of the literature. Subsequently, a validation study used the opinions of experts to confirm the numerical codes assigned to the responses. It was also necessary to test the items in each of the scales to see whether the model of a unidimensional scale was consistent with recorded data. It was possible to show that by using the numerical codes, the chosen items fitted well their respective scales. Once the three scales (a) effects of Society on Science and Technology (Society), (b) the effects of Science and Technology upon Society (Science), and (c) characteristics of Scientists (Scientists), had been specified and items were identified that satisfied the requirement of unidimensionality, it was possible to calibrate the three scales and the items within them using the partial credit model for Rasch scaling. The construction and calibration of these three scales permitted an investigation to proceed that involved the accurate measurement of students’, teachers’ and scientists’ views on STS.

Science, technology and society, Science education, Australia, Rasch scaling, scales, validation

PROBLEM

One of the aims of this Australian study was to develop a master scale to measure views, beliefs and attitudes towards Science, Technology and Society (STS). Oppenheim (1992) has stressed very strongly that a great deal of careful questioning, thought and consideration, and repeated conceptualisations, are necessary to produce effective attitude scales. Furthermore, he has suggested that after trialling in a pilot study a large number of attitude statements, this item pool should be analysed and submitted to a scaling procedure. The resultant scales, each of which would consist of a smaller number of statements than in the original item pool, could be used to allocate a numerical score for each respondent. A procedure similar to the one described by Oppenheim for the development of attitude scales was used in this study, although the study sought to construct what may best be considered as ‘view’ or ‘descriptive’ scales (Morgenstern and Keeves, 1997) rather than attitude scales.
The development of the scales, the calibration of the scales and the determination of how well the data from the respondents fit the requirements of Rasch scaling are discussed in this paper. Moreover, the validation of the scales by using the opinions of the experts is considered.

THE ORIGINAL VOSTS INSTRUMENT

The scales used in this Australian study were constructed and calibrated from the Views on Science-Technology-Society (VOSTS) instrument (Aikenhead, Fleming and Ryan, 1987) originally developed in Canada. The three scales finally used in this Australian study related to:

(1) effects of Society on Science and Technology (Society);
(2) effects of Science and Technology upon Society (Science); and
(3) characteristics of Scientists (Scientists).

Each item used in the scales addressed a particular issue. The item and the issue addressed by each particular item are shown in Appendix 1.

Aikenhead and Ryan (1992) believed that the processes used to write the VOSTS items from students' perspectives conferred an inherent validity to these items that were prepared from students’ actual statements. They believed that it was, therefore, not appropriate to speak in the traditional sense about the validity of VOSTS items. Moreover, Aikenhead and Ryan concluded that “the field of Item Response Theory has not yet developed the mathematical procedure that could analyse responses to VOSTS items” (Aikenhead and Ryan, 1992, p. 488).

It must be argued, nevertheless, that Aikenhead and Ryan were apparently unaware of the ways in which the partial credit model of item response theory might be employed to calibrate the students' responses to the items that they had developed, provided the items and the assigned code values were consistent with the requirement of unidimensionality associated with the scale domain that they had defined. In order to measure students' views and to compare their views with those of their teachers and professional scientists in a systematic way this present study challenged Aikenhead and Ryan's contention. In addition, this study sought to show that there was internal consistency within both the categories of items that Aikenhead and Ryan had assembled, and within the underlying views of students in relation to STS issues that provided strong meaning to their responses. Thus, the validation of the instrument was a necessary and critical component of this use of scale scores to compare the views of students, teachers and scientists, alongside the use of calibration procedures based on item response theory and the Rasch model.

MODELS OF MEASUREMENT

For the purposes of this Australian study, the form of statistical analysis used to establish unidimensionality and consequent interval scale measurement was the Rasch partial credit model (Rasch, 1960) based on the item response theory analysis of the students’ levels of performance in relation to the items of the instrument. This technique employs a latent trait model, and a theoretical approach to educational measurement with items in which more than two ordered levels of outcome are defined (Masters, 1988). This models has considerable advantages when compared with the classical test theory models. The characteristics and advantages of this model are discussed below, following a discussion of the limitations of classical test theory models.
Item response theory

The measurement processes in this Australian study employ a scaling model in which scale values are assigned to both the items and the respondents. As Keeves (1992) wrote in regard to item response theory, it is a requirement that there is a common underlying trait of performance for both respondents and the items which are used. The items and the respondents are located at levels along a scale defined by the latent trait. In this study, the latent trait is the strength and coherence of students’ views towards issues resulting from the relationships between science, technology and society. The position of an individual respondent on the latent trait scale is the level at which the respondent would answer, with a specified degree of probability (usually 0.50), an item located at that level on the scale (Keeves, 1992). In the scaling model used in this study, the respondents depend on the strength and coherence of their underlying views towards STS to respond to the items in a favourable manner.

The model of measurement used in this study is the partial credit model of item response theory. The two basic postulates of item response theory are:

1. the performance of an examinee on items can be predicted by factors called latent traits; and
2. the relationship between item performance and the set of underlying traits can be described by a monotonically increasing curve called an ‘item characteristic curve’ (Hambleton, Swaminathan and Rogers, 1991; Weiss and Yoes, 1991).

This item characteristic curve relates the probability of success on an item to the performance measured by the test and the characteristics of the item (Hambleton, Swaminathan and Rogers, 1991). The particular item response model used in this study is the Rasch partial credit model. The Rasch model is a one-parameter logistic model which has strong measurement properties. The advantage of item response theory over classical test theory is that the scores provided are not dependent on the specific set of items administered or the specific group of persons used in calibration (Weiss and Yoes, 1991).

In order to use the Rasch partial credit model, it is necessary to assume that the items occupy a space along the latent trait continuum that is consistent across both the entire group of students and the set of items being used (Keeves, 1988). Thus the partial credit model is a measurement model, since it provides a probabilistic connection between the categories of observed outcome on an item and the respondents' location on a latent trait of developing views. This probabilistic connection provides a basis for constructing measures of the respondent's views from a set of items with multiple outcome categories (Masters, 1988). Aikenhead and Ryan (1992) had not realised that item response theory had advanced to the point of encompassing several levels of student response. Hence this study, which uses item response theory to establish the consistency of the scaling of the VOSTS items developed by Aikenhead and his colleagues (Aikenhead, Fleming and Ryan, 1987) is a significant advance in research in this area.

Issues of consistency and validity

The first step in scaling the VOSTS items involved determining response values, or establishing which of the responses for each item should be allocated the chosen range of codes of 4, 3, 2, 1, or 0. This initial scaling was performed after consideration of the scholarly writing on science, technology and society. It was then important to examine the internal consistency of each scale using item response theory, and to establish the validity of each scale using the views of a panel of experts.
Through the use of the QUEST program (Adams and Khoo, 1993), the values for strength and coherence of the views of the respondents corresponding to each score level were determined and estimates were obtained for the difficulty parameters of each of the items in the scales. It was then possible to determine how well the Rasch model fitted the data. This was necessary, since as Wright (1988) has stressed, if the model did not fit the data, then it could not be used to calibrate items or to measure persons. Thus, in Wright's opinion, it was necessary for measurement to examine the validity of both item response patterns, and person response patterns, by evaluating the fit between the model and the data.

**Response validation by analysis of fit between the model and the data**

The codes for the different statements that comprised the items were specified and the location of each item estimated indicate its position along the graph of the latent trait generally, with the high codes at the top of the graph and the lower codes towards the bottom. The map would also show the position of the response groups that were formed on the same scale. It was expected that statements for each item with equivalent code values would be located at similar threshold levels along the latent trait axis. Moreover, if the codes for an item differed substantially in position from that of the equivalent codes for other items, either the item should be rejected, or the assigned codes should be reconsidered and adjusted. The scales used in the study could, in this way, be calibrated for use. If the fit of an item to the scale had an acceptable value, the item was considered to be consistent with the underlying scale.

The consistency of each scale was tested by examining the data gathered during a pilot study using a group of respondents with similar characteristics to those for whom the scales were intended (Hambleton and Zaal, 1991). In this study, the consistencies of the items and statements and the students' responses were initially examined before the scales were employed to measure the views of senior secondary science students in the 29 South Australian schools and colleges which were sampled for the main study. The internal consistency was therefore demonstrated, in the first instance, by how well the data fitted the Rasch scale. In this study, the analysis also demonstrated the degree to which the items satisfied the requirement of unidimensionality, which reflected the validity of the scale.

Furthermore, invariance of item and performance parameters, which only holds when the data fit the model well, is an indispensable feature of Rasch measurement, since it enables applications of the modelling procedures in such investigations as those of item bias and readily permits both vertical and horizontal equating (Hambleton, Swaminathan and Rogers, 1991). Therefore, it is very important to determine whether the partial credit model fitted the data in this study. The fit of the models to the student data was sufficiently strong for it to be argued that the scales were well-supported, and that the researcher could proceed confidently with the investigation and further data analysis.

**Internal consistency and validity in item response theory**

In evaluating the instrument used in this Australian study, it was important to address the issues of validity as well as internal consistency and reliability, in order to determine the generalisability of the scores allocated to respondents, and the strength of measurement that could be made using the scores. After the internal consistency of the items was examined using Rasch scaling the validity was considered. One conception of validity is how faithfully the set of items in an instrument correspond to that attribute in which the researchers are interested (Hambleton and Rogers, 1991). The scores are valid if, in fact, the instrument is seen to measure what it purports to measure.
In order to test the validity of each scale seven persons who were recognised scholars in the field of Science, Technology and Society (STS) were asked to rank the statements in each item on a five-point scale according to their understanding of the underlying STS dimension. These aggregated rankings could then be compared with the scale scores assigned by the researchers from a systematic review of the literature on STS.

Table 1 for Item No. 3 shows the results for the assigned scores (Scale scores) for this item in comparison with the experts' responses that are recorded as mean scores (Expert mean scores), ranks (Expert ranks) and integral scale scores (Expert scores). After the experts' scores were collected, the items were separated by scale, and three tables were prepared, with the items in presentation order. The agreement between the experts' scores and the final scores used in this present study could then be examined. First, the experts' scores were pooled and the experts' mean scores for the responses for each item were calculated. The experts' mean scores were then ranked. Overall experts' scores were subsequently determined according to this ranking.

Table 1. Validation of response scores using the ratings of the experts - Scale: Society

<table>
<thead>
<tr>
<th>Item No. 3</th>
<th>Response Category</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expert mean scores</td>
<td>2.00</td>
<td>2.43</td>
<td>3.14</td>
<td>3.00</td>
<td>1.43</td>
<td>1.86</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Expert ranks</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Expert scores</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Scale scores</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>p</td>
<td>0.14</td>
<td>a=</td>
<td>0.94</td>
<td>% ag</td>
<td>c=</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| a: probability value; b: reliability coefficient; c: percentage agreement

Secondly, the level of agreement, and significance of any differences between the researcher's scale scores and overall experts' scores were examined using $\chi^2$ values and percentage agreement. For each item, the $\chi^2$ value and its probability were calculated using the reliability analysis routine in SPSS (Version 7) with the Friedman ANOVA option, with zero forming only one category. The results for each scale are given in Table 2.

The reliability coefficient was obtained as a Cronbach $\alpha$ coefficient also from the SPSS program. The percentage agreement was calculated for each item by dividing the number of responses where there was complete agreement between the researcher's scale scores and the experts' scale scores by the total number of responses for the item. For Item 3, by way of example, the percentage agreement (0.57) is calculated by dividing the number of responses which are in agreement (4) by the total number of responses (7). Where there is a very high level of agreement, the reliability coefficient goes to one, and the per cent agreement goes to one.

The coefficients of concordance were not employed to indicate the significance of any difference between the scores and the experts' score, since they could not be used for all cases. However, the method of validation using $\chi^2$ values, reliability coefficients and percentage agreement showed that while the experts' scores and those of the researcher were not in complete agreement, there was sufficient agreement to indicate that the scoring employed was valid or strong in so far as it was in agreement with the scoring provided by the experts.
Table 2. Summary statistics for the validation of response scores using the ratings of the experts for Society, Science and Scientists Scales

<table>
<thead>
<tr>
<th>Item No.</th>
<th>$\chi^2_{a}$</th>
<th>$p^a$</th>
<th>$R C^b$</th>
<th>% ag $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Society Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\chi^2 = 4.0$</td>
<td>0.14</td>
<td>0.94</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>$\chi^2 = 3.0$</td>
<td>0.39</td>
<td>0.92</td>
<td>0.44</td>
</tr>
<tr>
<td>9</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>$\chi^2 = 3.2$</td>
<td>0.20</td>
<td>0.93</td>
<td>0.67</td>
</tr>
<tr>
<td>15</td>
<td>zero</td>
<td></td>
<td>0.98</td>
<td>0.86</td>
</tr>
<tr>
<td>19</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>21</td>
<td>$\chi^2 = 0.0$</td>
<td>1.0</td>
<td>0.93</td>
<td>0.50</td>
</tr>
<tr>
<td>23</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>25</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Science Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\chi^2 = 4.4$</td>
<td>0.35</td>
<td>0.85</td>
<td>0.38</td>
</tr>
<tr>
<td>2</td>
<td>$\chi^2 = 3.2$</td>
<td>0.16</td>
<td>0.97</td>
<td>0.67</td>
</tr>
<tr>
<td>7</td>
<td>zero</td>
<td></td>
<td>0.98</td>
<td>0.86</td>
</tr>
<tr>
<td>10</td>
<td>$\chi^2 = 2.0$</td>
<td>0.16</td>
<td>0.97</td>
<td>0.78</td>
</tr>
<tr>
<td>14</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>16</td>
<td>$\chi^2 = 2.0$</td>
<td>0.37</td>
<td>0.93</td>
<td>0.63</td>
</tr>
<tr>
<td>17</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>zero</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>24</td>
<td>zero</td>
<td></td>
<td>0.95</td>
<td>0.71</td>
</tr>
<tr>
<td>26</td>
<td>zero</td>
<td></td>
<td>0.89</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>Scientists Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>zero</td>
<td></td>
<td>0.99</td>
<td>0.90</td>
</tr>
<tr>
<td>6</td>
<td>zero</td>
<td></td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>zero</td>
<td></td>
<td>0.95</td>
<td>0.60</td>
</tr>
<tr>
<td>11</td>
<td>$\chi^2 = 3.2$</td>
<td>0.20</td>
<td>0.94</td>
<td>0.50</td>
</tr>
<tr>
<td>13</td>
<td>$\chi^2 = 2.0$</td>
<td>0.16</td>
<td>0.97</td>
<td>0.78</td>
</tr>
<tr>
<td>18</td>
<td>$\chi^2 = 5.8$</td>
<td>0.21</td>
<td>0.89</td>
<td>0.28</td>
</tr>
<tr>
<td>20</td>
<td>$\chi^2 = 5.3$</td>
<td>0.38</td>
<td>0.89</td>
<td>0.40</td>
</tr>
<tr>
<td>27</td>
<td>$\chi^2 = 4.4$</td>
<td>0.22</td>
<td>0.94</td>
<td>0.56</td>
</tr>
</tbody>
</table>

a: probability value; b: reliability coefficient; c: percentage agreement

The Calibration of the Scales

During this study, analysis of the fit of a unidimensional model to the data was conducted by using Rasch scaling to place the people and the items on the same unidimensional interval scale. The fit between the model and the data was established primarily by using the item infit mean square values. This analysis also examined the item analysis results obtained with classical test theory,
item difficulty level estimates or threshold values and maps of respondents and items for each scale. Hence, it is important at this stage to describe, in some detail, the techniques of analysis which were employed.

In item response theory, the probability of a correct response depends on both the respondent’s performance level and on the difficulty parameter of the item, which is its difficulty level or threshold value. In this study, both the respondents’ performances and the item threshold parameters were initially unknown. After the data had been collected and entered, the performance parameters of respondents and the Thurstone threshold values for the response categories for each item and the fit parameters of the items and the persons were estimated using the QUEST computer program (Adams and Khoo, 1993). The parameters that were associated with the theoretical item response curve were estimated using a maximum likelihood procedure.

A minor disadvantage of this procedure was that performance estimates associated with perfect and zero scores did not provide information for the analysis. There were no zero scores in the sample, so these did not need to be considered. It was necessary, however, to exclude those students obtaining perfect scores in the main analysis and subsequently to calculate the estimates for their perfect scores. The units on this scale are logits (Hambleton and Swaminathan, 1985). The following section considers the information provided in the QUEST program for the analysis of a particular item. Appendix 2 records the items included in the three scales.

**Item Analysis and Response Values**

Table 3 shows the item analysis results for observed responses to Item 3 for students for the three scales. There are nine possible multiple-choice responses for Item 3, and ten categories of response: 0,1,2,3,4,5,6,7,8, and 9 if the omitted response is also considered. The QUEST computer program (Adams and Khoo, 1993) was used for item analysis in this study. The first task in the development of a scoring procedure for the VOSTS instrument involved the recoding from the original alphabetical or numerical coding system for the responses for each item to item characteristic scores, between zero and four, which were allocated to the responses for particular items. Hence, these numerical categories of response were recoded to 0,2,2,3,4,1,1,0,0, and 0 once scores were allocated to each response. This formed a five-point scale, since scores from 0-4 were used. The initial scores were assigned by the researcher based on an understanding of STS theoretical positions. Those response categories most consistent with the 'ideal' STS position, were assigned a score of 4 and those least consistent with the 'ideal' STS position were assigned a score of 0, with intermediate scores assigned accordingly. The item analysis information for each of the categories includes: counts, percentages, point-biserial correlations, p-values, mean ability levels, step labels, Thurstone thresholds and errors.

This procedure was similar for all of the other items, and the results for the observed responses demanded very careful examination during this stage of the study. A substantial amount of very important information was provided by the distractors or alternative response categories that Aikenhead and his colleagues developed, and it was important to use this information, rather than to ignore it.

The point-biserial values (point-biserial correlations) provide information on the discrimination for each response statement (Hambleton and Swaminathan, 1985). For Item 3 (see Table 3), the point-biserials related well with the mean performance values for students responding to this response statement.
Table 3. Item Analysis results for observed responses: students on Item 3 on Society Scale (N = 1278; L = 9)

<table>
<thead>
<tr>
<th>Categc</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Code</td>
<td>203</td>
<td>109</td>
<td>147</td>
<td>382</td>
<td>277</td>
<td>83</td>
<td>14</td>
<td>32</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Countsd</td>
<td>3</td>
<td>203</td>
<td>109</td>
<td>147</td>
<td>382</td>
<td>277</td>
<td>83</td>
<td>14</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>Percente</td>
<td>0.2</td>
<td>15.9</td>
<td>8.5</td>
<td>11.5</td>
<td>29.9</td>
<td>21.7</td>
<td>6.5</td>
<td>1.1</td>
<td>2.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Pt-Biserf</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.07</td>
<td>0.07</td>
<td>0.37</td>
<td>-0.13</td>
<td>-0.17</td>
<td>-0.21</td>
<td>-0.23</td>
<td>-0.15</td>
</tr>
<tr>
<td>p-valueg</td>
<td>.221</td>
<td>.425</td>
<td>.005</td>
<td>.004</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Mean Abilityh</td>
<td>0.02</td>
<td>0.23</td>
<td>0.11</td>
<td>0.35</td>
<td>0.53</td>
<td>0.10</td>
<td>-0.10</td>
<td>-0.84</td>
<td>-0.61</td>
<td>-0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step Labels</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thresholdi</td>
<td>-1.75</td>
<td>-0.13</td>
<td>0.41</td>
<td>0.73</td>
</tr>
<tr>
<td>Error</td>
<td>0.16</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
</tr>
</tbody>
</table>

a: discrimination index;  b: infit mean square;  c: response category; d: number of respondents; e: percentage of respondents; f: point-biserial correlations; g: probability that the point biserial correlations are significantly different from zero; h: mean ability level for respondents; i: Thurstone thresholds or difficulty levels

By way of example, response Category 4, which drew 382 (29.9%) students has been assigned the highest score of four and has the highest mean ability level for respondents of 0.53 with a point-biserial correlation of 0.37. The next highest score is three for Category 3. The mean ability level is 0.35, resulting in the second highest mean ability level for all response categories and a point-biserial correlation of 0.07. Some of the lowest point-biserial values are for categories 0, 7, 8, and 9 with negative point-biserial correlations of -0.02, -0.21, -0.23, and -0.15 respectively. These categories also have low mean respondent ability levels of 0.02, -0.84, -0.61, and -0.30 respectively.

The p-values merely indicate the probability that the point-biserial correlations are significantly different from zero. The mean ability values in the item analysis also provide important information since the less favourable the response category towards the 'ideal' STS position, the lower the mean ability value. The mean ability values indicate the relative degree of strength and coherence of a response category, with the stronger and more coherent response categories having the highest mean ability values. Hence, the mean ability levels for Category 4 (0.53) and Category 3 (0.35) indicate response categories closest to the 'ideal' STS position.

In the pilot study, in those few cases where discrepancies were observed, some adjustments were made since it was considered important to re-consider the scaling. The students were found, overall, to have quite strong, coherent and well-directed views towards STS. The review of the literature was taken into account during this process of further reflection upon the scores allocated to each of the responses.

The overall consistency between the scaled response categories and the total scores of the students on the scale is given by the item discrimination index. In general, these indices are moderate or strong and the value of 0.47 is recorded for Item 3. Table 3 shows that for Item 3, the calculated value of the infit mean square deviated only slightly (1.06) from the expected value of 1.00.

**Item Threshold Estimates**

Table 4 presents the item estimates and threshold values for students’ responses to the Society items scale showing the thresholds (between 1 and 4), and the infit and outfit mean square values and t-values for each item. Table 4 also shows the score for the item using score values (0 to 4) and the maximum overall score. The outfit mean square values give the overall fit of the item to the
scale across all positions on the scale, and all persons in the sample. The infit mean square value is an index of item fit at the steepest part of the item characteristic curve. The fit of the item to the unidimensional scale is thus given by the infit mean square, with an expected value of 1.00 for an item that discriminates well. The infit mean square values should lie within the range of 0.83 to 1.20 for the item to be considered to fit the model well. In the analysis of the trial data, one item with infit mean square value of 0.80 was found to discriminate too well, and another item with an infit mean square value of 1.22 did not discriminate well enough.

The item threshold statistics show the difficulty levels associated with the response categories of an item (Masters, 1988). Table 4 shows that for Item 3, the transition levels at the 0.5 probability level of response from the score of 0 to 1, from 1 to 2, from 2 to 3, and from 3 to 4 are given by the threshold levels of -1.75, -0.33, 0.37, and 0.73 respectively. These threshold values form a clear monotonic sequence with relatively small errors (approximately 0.10) involved in estimation. The threshold for a score of four is 0.73, which is higher than the threshold of 0.41 for a score of three (see Table 4). The threshold for a score of one is -1.75, that is substantially lower than the threshold for the higher scores. Consequently, the fit estimates for this item conform well with the requirements of the model. The infit and outfit mean square values and t-values for this item are also within the range (approximately one) which indicate that the item discriminates consistently with respect to the other items in the scale.

However, the outfit mean square value is primarily used to assess the fit of persons to the model, and the t-statistics that are heavily influenced by sample size and in this study with 1278 students, are not taken into consideration. Nevertheless, it should be noted that for this scale both the infit and outfit t-values are less than the critical level of 2.0 for non-fit of an item to the scale. The fit data for all the items are presented in Appendix 3.

Table 4. Item estimates (thresholds) in input order: students on Society Scale (N = 1278; L = 9)

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Score</th>
<th>Maxa</th>
<th>Thurstone Thresholds</th>
<th>Inftb Mnsq</th>
<th>Outftc Mnsq</th>
<th>Inftd t</th>
<th>Outfte t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 3</td>
<td>2949</td>
<td>5112</td>
<td>1 -1.75 -0.13 0.41 0.73</td>
<td>1.06 1.08 1.8 1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td>2613</td>
<td>5112</td>
<td>.16 .09 .09 .10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td>3125</td>
<td>5112</td>
<td>.16 .11 .11 .10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 12</td>
<td>2819</td>
<td>5112</td>
<td>.11 .12 .09 .10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 15</td>
<td>2829</td>
<td>5112</td>
<td>.09 .09 .09 .11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 19</td>
<td>3213</td>
<td>5112</td>
<td>.09 .09 .10 .10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 21</td>
<td>2943</td>
<td>5112</td>
<td>.11 .11 .10 .09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 23</td>
<td>3115</td>
<td>5112</td>
<td>.09 .09 .08 .08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 25</td>
<td>2543</td>
<td>5112</td>
<td>.09 .09 .08 .11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>1.01</td>
<td>1.03 0.2 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.13</td>
<td>0.04</td>
<td>0.06 1.2 1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a: maximum score  b: infit mean square values  c: outfit mean square values  d: infit t values  e: outfit t values
Calculating Perfect Scale Scores

Table 5 shows the score equivalents for all score values for the *Society* scale. Since the computer program did not calculate perfect scale scores, it was necessary to estimate the score for students on the three scales who gained a maximum score of 100 per cent. This was accomplished by consulting the score equivalence tables for the three scales and using the top three logits for calculations.

<table>
<thead>
<tr>
<th>Score</th>
<th>Estimate (logits)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>max = 36</td>
<td>0 (estimate)</td>
<td>3.34</td>
</tr>
<tr>
<td>35</td>
<td>2.51</td>
<td>0.91</td>
</tr>
<tr>
<td>34</td>
<td>1.93</td>
<td>0.64</td>
</tr>
<tr>
<td>33</td>
<td>1.60</td>
<td>0.51</td>
</tr>
<tr>
<td>32</td>
<td>1.37</td>
<td>0.45</td>
</tr>
<tr>
<td>31</td>
<td>1.19</td>
<td>0.39</td>
</tr>
<tr>
<td>30</td>
<td>1.04</td>
<td>0.36</td>
</tr>
<tr>
<td>29</td>
<td>0.92</td>
<td>0.34</td>
</tr>
<tr>
<td>28</td>
<td>0.81</td>
<td>0.32</td>
</tr>
<tr>
<td>27</td>
<td>0.71</td>
<td>0.31</td>
</tr>
<tr>
<td>26</td>
<td>0.62</td>
<td>0.29</td>
</tr>
<tr>
<td>25</td>
<td>0.53</td>
<td>0.28</td>
</tr>
<tr>
<td>24</td>
<td>0.46</td>
<td>0.27</td>
</tr>
<tr>
<td>23</td>
<td>0.38</td>
<td>0.27</td>
</tr>
<tr>
<td>22</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>21</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>20</td>
<td>0.17</td>
<td>0.26</td>
</tr>
<tr>
<td>19</td>
<td>0.10</td>
<td>0.26</td>
</tr>
<tr>
<td>18</td>
<td>0.04</td>
<td>0.26</td>
</tr>
<tr>
<td>17</td>
<td>-0.03</td>
<td>0.26</td>
</tr>
<tr>
<td>16</td>
<td>-0.10</td>
<td>0.26</td>
</tr>
<tr>
<td>15</td>
<td>-0.17</td>
<td>0.26</td>
</tr>
<tr>
<td>14</td>
<td>-0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>13</td>
<td>-0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>12</td>
<td>-0.39</td>
<td>0.28</td>
</tr>
<tr>
<td>11</td>
<td>-0.47</td>
<td>0.29</td>
</tr>
<tr>
<td>10</td>
<td>-0.56</td>
<td>0.30</td>
</tr>
<tr>
<td>9</td>
<td>-0.65</td>
<td>0.32</td>
</tr>
<tr>
<td>8</td>
<td>-0.76</td>
<td>0.33</td>
</tr>
<tr>
<td>7</td>
<td>-0.88</td>
<td>0.36</td>
</tr>
<tr>
<td>6</td>
<td>-1.02</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>-1.19</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>-1.40</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>-1.67</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>-2.08</td>
<td>0.71</td>
</tr>
<tr>
<td>1</td>
<td>-2.78</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Table 5 shows that the maximum score is 36. The logits for scores of 33, 34, and 35 are 1.60, 1.93 and 2.51 respectively. The maximum logit is calculated by adding to the logit for highest score (35) shown on the score equivalence table (2.51) the difference between 2.51 and 1.93 (0.58) and the difference between 0.58 and 0.33 (1.93 minus 1.60). The figure obtained from subtracting 0.33 from 0.58 is 0.25. Thus, the maximum logit = 2.51+ 0.58+ 0.25= 3.34.

Maps of Respondents and Item Thresholds

The map of respondents and item thresholds for the Society scale presented in Figure 1 shows the location of the item thresholds and the respondents on the scale. The clustering of the item levels provides solid evidence for the internal consistency and validity of both the model and the scores obtained as a result of the study. Both person and item locations are shown in the maps for the three scales so they are extremely useful for the subsequent analysis of observed responses. There is more than one threshold per item on these maps because it has employed the partial credit model. The locations of the item thresholds in Figure 1 show that the levels associated with thresholds for Level 4 in the main are above the levels of thresholds for Level 3, which are higher on the scale than thresholds for Level 2. This further confirms the consistency of the scores assigned to the responses to the items in the instrument used in this study. Furthermore, the map also shows that the item thresholds are not equally spaced, but are located at particular levels on the latent trait scale. Likewise the persons are located at particular levels on the latent trait scale. The higher the level of the person on the y-axis, the greater the probability that he or she will give a strong and coherent response to an item. Similarly, the higher the level of an item threshold on the y-axis, the greater the strength and coherence of this item towards STS.

The item responses and numbers of respondents for Item 3 are shown on the item estimates map for all students on the Society Scale. Although the response thresholds for this item are lower than for many of the other items, the thresholds are well placed in comparison with the thresholds for the other items.

COMPARISON WITH OTHER METHODS USED TO VALIDATE ATTITUDE SCALES

There are many similarities between the methods of validation used for the attitude scales developed previously by other researchers (Rubba, Bradford and Harkness, 1996) and the validation of the scales developed in this present study. Whereas in the present study, the experts' scaling followed the scaling of the items based on a review of the literature and the researchers’ judgment, in the study by Rubba, Bradford and Harkness, the scaling was performed exclusively by the panellists. In this present study, techniques discussed in this paper were used to establish the agreement between the experts' scores and the scale scores allocated by the researchers on the basis of the review of the literature.

Although Rubba and his colleagues scaled the VOSTS instrument in their study, and used the views of experts to validate the scales, as has been accomplished in this Australian study, they did not examine the items and the response categories in the way that has been done in this study. These American researchers just assumed that the experts had given the correct responses. In this Australian study it was considered important to calibrate the scaled information and the procedures to establish the validity and consistency of the scales.
<table>
<thead>
<tr>
<th>2.0</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1.0</th>
<th>xxx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>xxx</td>
</tr>
<tr>
<td></td>
<td>xxx</td>
</tr>
<tr>
<td></td>
<td>xxx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0.0</th>
<th>xxxxxxx</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>xxxxxx</td>
</tr>
<tr>
<td></td>
<td>xxxx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-1.0</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

| -2.0 | 5.1 |

<table>
<thead>
<tr>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>xxx</td>
</tr>
</tbody>
</table>

| Scale zero set at mean of item thresholds |

Each x represents 10 students

Figure 1. Map of respondents and item estimates (thresholds): all students on Society Scale (N = 1278; L = 9)
Consideration of the validity of attitude scales used in previous studies further highlights the appropriateness of the methods which were used in this present study to establish the validity of the items in the scales. The Scientific Attitude Inventory (SAI) (Moore and Sutman, 1970) was revised by Moore and Foy (1997) in order to improve readability and remove gender-biased language. There had been criticism of the validity of the original instrument (Munby, 1983). This criticism was even more pronounced in regard to the revised instrument (Munby, 1997) and was quite important and worthy of consideration. During the process of the revision of the SAI, the researchers argued that the validity of the original instrument, which was established using the judgments of a panel of judges, was maintained in the final SAI II instrument, since the position of statements in the instrument had not changed. The authors wrote that:

since there is evidence for the content validity of the items in the original instrument with respect to the 12 position statements, we decided to make as few changes as possible while responding to criticisms and suggestions. This evidence was presented in item selection and the field test of the original SAI. (Moore and Foy, 1997, p. 329)

In the original instrument, attitude statements were selected for use from a pool of items after the judges judged each attitude statement in terms of whether it represented a particular position statement (Moore and Sutman, 1970). This method of validation differs substantially from the method in the present study. Furthermore, the panel of seven experts from the Australasian Association for the History, Philosophy and Social Studies of Science provided independent scoring of the items in the scales, and the agreement between the experts’ scale scores and those of the researchers was established using well-recognised statistical techniques.

The construct validity of the original SAI was demonstrated in a field test (Moore and Sutman, 1970). During this test, the SAI was administered to three groups of low-ability tenth-grade biology students. The investigator presented lessons to each of these groups. The series of lessons for the first group, the control group, was the regular sequence prepared by the teacher of this group. The other two groups received lessons which were specially designed to develop the attitudes assessed by the SAI. The authors believed that since both of the second two groups who received instruction relevant to the development of the scientific attitudes in the SAI had significantly higher post-test means than the post-test means of the control group, this field test showed that the SAI had construct validity. Munby (1983) questioned the method by which Moore and Sutman (1970) believed they had validated the SAI instrument. He argued that it was not certain what was measured by the SAI, and many of the items in the SAI, that were believed to gauge attitudes, could be interpreted very differently. Following his study of the SAI, Munby contended that "there are sufficient grounds for judging the SAI to be conceptually doubtful if not weak” (Munby, 1983, p. 157).

After their redevelopment of the SAI, Moore and Foy (1997) argued that they sought to show the validity of SAI II with a confirmatory factor analysis of data from 557 respondents. However, this resulted in a reduced number of items in five scales, and the grouping of these items was not very satisfactory in providing meanings for the item groups. The authors concluded that:

regrouping the items would virtually eliminate the support gained by judges for the validity of the instrument. Therefore, the 40-item SAI II is being advanced as presented here without the support of factor analysis. (Moore and Foy, 1997, p. 332)

These findings cast doubt upon Moore and Foy's assertion in regard to their redeveloped SAI II scale that:

it is possible that the objects of the scales are so ill-formed in the subjects, students in this case, that we are not able to use their responses to confirm the scales as factors. (Moore and Foy, 1987, p. 333)
Munby (1997) was rightfully very critical of Moore and Foy's claims in regard to the validity of the SAI II instrument. He showed that empirical work with the SAI II raised doubts about its validity. Munby decried the statement by Moore and Foy about the objects of the scales being so ill-formed that it was not possible to use students’ responses to confirm the scales as factors. Munby argued that;

This statement suggests that the authors put more credence in the evidence of validity obtained from the panel of judges using the older version of the SAI than they do in the empirical determination using the present version. (Munby, 1997, p. 338)

Both the empirical methods and the validation by a panel of experts which were used in the development of the scales employed in the present study have been discussed in detail. Moreover, the validation by a panel of experts differed substantially from that used by Moore and Foy. The examination of the scores assigned by the researchers and the experts employed statistical techniques to show the validity of the scoring of the scales. The need to establish the way in which the empirically-derived responses corresponded with the scales was supported by Munby (1996).

Once the fit of the student information to the scale to establish the consistency of the scoring of the scales, and the validity of the scales had been ensured, an investigator could proceed with confidence to measure respondents’ views using these scales.

**SUMMARY AND CONCLUSION**

The scales used in this present study were developed from items in the VOSTS inventory, with careful consideration of the nature of the items used in the final instrument, as well as the consistency and the validity of the scales. As a first step in the development of the scales, each of the response categories was assigned a numerical score based on a judgment of its degree of consistency with the ‘ideal’ STS view. The final instrument was prepared after the examination of the data collected during the trial study.

A detailed literature review was an important form of validation of the scale values assigned to the responses for the items in the instrument. Furthermore, the establishment of the validity of the scales by comparison with the views of the experts is presented in this paper.

The trial data were used to examine whether the items discriminated sufficiently between respondents who were high on the scale and those who were low. The numbers of students who selected particular scaled responses were shown on maps for each scale, where the vertical axis represented the overall levels of strength and coherence towards STS of respondents' views. The QUEST program estimated the thresholds where one scaled response changed to another and these Thurstone thresholds have been presented. It was expected that most of the very strong and coherent scores (fours), would be at the top of the graph. This was found to occur.

Consequently, after the data for the pilot study were entered and processed, the consistency of each item to its scales was analysed, since the analysis provided evidence for a test of goodness of fit for individual items to the particular scales. It was possible to use Rasch scaling to investigate the fit of the model to the data, since the Rasch scaling model places people and items on the same interval scale. Infit mean square values were used to establish the fit between the model and the data. Items that did not conform to the model were eliminated, due to their inability to differentiate the strength and coherence of respondents' views towards STS in a manner that was consistent with the other items. This analysis of the data also examined the item analysis results for observed responses, item difficulty level estimates or threshold values and maps of respondents and items for each scale.
In conclusion, the partial credit model and Rasch measurement procedures provide a very powerful approach to the scaling of views and attitudes which sets the scores obtained on an interval scale that is not only independent of the items and statements employed, but also the sample of persons used to calibrate the scale.

REFERENCES


APPENDIX 1: ISSUES ADDRESSED BY THE ITEMS IN THE SCALES

The Effects of Society on Science and Technology (Society)
- The Influence of Governments on Science (Items 3, 9, 19, and 25)
- The Influence of the Military on Science. Case History: Scientists in the Second World War (Item 23)
- The Influence of Special Interest Groups on Science (Item 12)
- The Influence of Educational Establishments on Science (Items 5 and 21)
- The Control of Science by Corporations (Item 15)

The Effects of Science and Technology upon Society (Science)
- Social Responsibility of Scientists (Item 10)
- Science for the Solution of Practical Problems in Everyday Life (Item 1)
- Contributions to Social Decisions in Relation to Science and Technology (Item 16)
- The Contribution of Science and Technology to Economic Development (Items 2 and 22)
- Unwanted Effects Caused by the Application of Science (Item 17)
- The Effects of Capital-intensive Technology (Items 7 and 24).
- Science, Technology and the Social Fabric of Society (Items 14 and 26)

Characteristics of Scientists
- The Effect of the Personal Characteristics of Scientists on Science (Items 6 and 18)
- Australian Scientists’ Motivation for Doing Science (Item 27)
- The Effect of the Values of Scientists on Science (Item 8)
- Individual Scientists and the Scientific Method (Item 11)
- Under-representation of Females in Science (Item 13)
- Gender Effects on the Outcomes of Science (Items 4 and 20)

APPENDIX 2: ITEMS IN THE SCALES

Appendix 2.1 Item Number 1

In your everyday life, knowledge of science and technology helps you solve practical problems (for example, getting a car out of the sand, cooking, or caring for a pet).

Your position, basically: (Please read from A to J, and then choose ONE only from this page.)

The systematic reasoning taught in science classes (for example, hypothesising, gathering data, being logical):

A. helps me solve some problems in my daily life. Everyday problems are more easily and logically solved if treated like science problems. (2)
B. gives me greater knowledge and understanding of everyday problems. However, the problem solving techniques we learn are not directly useful in my daily life. (3)
C. Ideas and facts I learn from science classes sometimes help me solve problems or make decisions about such things as cooking, keeping healthy, or explaining a wide variety of physical events. (3)
D. The systematic reasoning and the ideas and facts I learn from science classes help me a lot. They help me solve certain problems and understand a wide variety of physical events (for example, thunder or quasars). (4)
E. What I learn from science class generally does not help me solve practical problems; but it does help me notice, relate to, and understand, the world around me. (2)

What I learn from science class does not relate to my everyday life:

F. Biology, chemistry and physics are not practical for me. They emphasise theoretical and technical details that have little to do with my day-to-day world. (1)
G. My problems are solved by past experience or by knowledge unrelated to science and technology. (2)
H. I don't understand. (0)
I. I don't know enough about this subject to make a choice. (0)
J. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis
Appendix 2.2 Item Number 2

The more Australia’s science and technology develop, the wealthier Australia will become. Your position, basically: (Please read from A to H and then choose ONE only from this page.)

Science and technology will increase Australia’s wealth:
- A. because science and technology bring greater efficiency, productivity and progress. (2)
- B. because more science and technology would make Australia less dependent on other countries. We could produce things for ourselves. (2)
- C. because Australia could sell new ideas and technology to other countries for profit. (3)
- D. It depends on which science and technologies we invest in. Some outcomes are risky. There may be other ways besides science and technology that create wealth for Australia. (4)
- E. Science and technology decrease Australia’s wealth because it costs a great deal of money to develop science and technology. (1)
- F. I don’t understand. (0)
- G. I don’t know enough about this subject to make a choice. (0)
- H. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.3 Item Number 3

The Australian government should give scientists research money to explore the curious unknowns of nature and the universe. Your position, basically: (Please read from A to I and then choose ONE only from this page.)

Money should be spent on scientific research:
- A. so Australia does not fall behind other countries and become dependent upon them. (2)
- B. in order to satisfy the human urge to know the unknown; that is, to satisfy scientific curiosity. (2)
- C. even though it’s often impossible to tell ahead of time whether the research will be beneficial or not. It’s an investment risk but we should take it. (3)
- D. because by understanding our world better scientists can make it a better place to live in (for example using nature’s environment and resources to our best advantage, and by inventing helpful technology). (4)
- E. only when the research is directly related to our health (especially finding cures for diseases), to our environment or to agriculture. (1)
- F. Little or no money should be spent on scientific research because the money could be spent on other things such as helping Australia’s unemployed and needy, or helping less fortunate countries. (1)
- G. I don’t understand. (0)
- H. I don’t know enough about this subject to make a choice. (0)
- I. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.4 Item Number 4

There are many more women scientists today than there used to be. This will make a difference to the scientific discoveries which are made. Scientific discoveries made by women will tend to be different from those made by men. Your position, basically: (Please read from A to M, and then choose ONE only from this page.)

There is NO difference between female and male scientists in the discoveries they make:
- A. because any good scientist will eventually make the same discovery as another good scientist. (2)
- B. because female and male scientists experience the same training. (2)
- C. because overall, women and men are equally intelligent. (3)
- D. because women and men are the same in terms of what they want to discover in science. (2)
- E. because research goals are set by demands or desires from others besides scientists. (3)
- F. because everyone is equal, no matter what they do. (1)
- G. because any differences in their discoveries are due to differences between individuals. Such differences have nothing to do with being male or female. (4)
- H. Women would make somewhat different discoveries because, by nature or by upbringing, females have different values, viewpoints, perspectives, or characteristics (such as sensitivity toward consequences). (4)
I. Men would make somewhat different discoveries because men are better at science than women. (1)
J. Women would likely make somewhat better discoveries than men, because women are generally better than men at some things such as instinct and memory. (1)
K. I don't understand. (0)
L. I don't know enough about this subject to make a choice. (0)
M. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.5 Item Number 5

The success of science and technology in Australia depends on us having good scientists, engineers and technicians. Therefore, Australia should require students to study more science in school. 

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

Students should be required to study more science:
A. because it is important for helping Australia to keep up with other countries. (1)
B. because science affects almost every aspect of society. As in the past, our future depends on good scientists and technologists. (3)
C. Students should be required to study more science, but a different kind of science course. Students should learn how science and technology affect their everyday lives. (4)

Students should NOT be required to study more science:
D. because other school subjects are equally or more important to Australia's successful future. (2)
E. because it won't work. Some people don't like science. If you force them to study it, it will be a waste of time and will turn people away from science. (2)
F. because not all students can understand science, even though it would help them in their life. (1)
G. because not all students can understand science. Science is not really necessary for everyone. (1)
H. because it's not right for someone else to decide if a student should take more science. (1)
I. I don't understand. (0)
J. I don't know enough about this subject to make a choice. (0)
K. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.6 Item Number 6

The best scientists have the patience and determination to get through times of frustration and boredom (for example, doing the same experiment many times to get reliable results). 

Your position, basically: (Please read from A to G, and then choose ONE only from this page.)
A. Yes, because frustration and boredom challenge the best scientist to struggle and work even harder. (1)
B. Yes, because patience and determination are part of the job. Without them, scientists would not get absolutely correct results. (4)
C. No, because even some of the best scientists cannot cope with frustration. Scientists have varying degrees of patience, like everyone else. (3)
D. No, because the best scientists are clever enough to avoid most frustration and boredom. Frustration and boredom make it harder for anyone to succeed. (2)
E. I don't understand. (0)
F. I don't know enough about this subject to make a choice. (0)
G. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.7 Item Number 7

Heavy industry has greatly polluted Europe and North America. Therefore, it is a responsible decision to move heavy industry to underdeveloped countries where pollution is not so widespread.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)
A. Heavy industry should be moved to underdeveloped countries to save developed countries and their future generations from pollution. (1)
B. It's hard to tell. By moving industry, developed countries would help poor countries to prosper, and developed countries would help reduce their own pollution. But they have no right to pollute someone else's environment. (2)
C. It doesn't matter where industry is located. The effects of pollution are global. (3)
Heavy industry should NOT be moved to underdeveloped countries:
D. because moving industry is not a responsible way of solving pollution. Developed countries should reduce or eliminate their own pollution, rather than create more problems elsewhere. (4)
E. because those countries have enough problems without the added problem of pollution. (2)
F. because pollution should be confined as much as possible. Spreading it around would only create more damage. (2)
G. I don't understand. (0)
H. I don't know enough about this subject to make a choice. (0)
I. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.8 Item Number 8
A scientist's religious views will NOT make a difference to the scientific discoveries he or she makes. Your position, basically: (Please read from A to G, and then choose ONE only from this page.)
A. Religious views do not make a difference. Scientists make discoveries based on scientific theories and experimental methods, not on religious beliefs. Religious beliefs are outside the domain of science. (2)
B. It depends on the particular religion itself, and on the strength or importance of an individual's religious views. (3)
Religious views do make a difference:
C. because religious views will determine how you judge scientific ideas. (1)
D. because sometimes religious views may affect what scientists do or what problems they choose to work on. (4)
E. I don't understand. (0)
F. I don't know enough about this subject to make a choice. (0)
G. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.9 Item Number 9
Community or government agencies should tell scientists what to investigate; otherwise scientists will investigate what is of interest only to them. Your position, basically: (Please read from A to J, and then choose ONE only from this page.)
Community or government agencies should tell scientists what to investigate:
A. so that the scientists' work can help improve society. (2)
B. only for important public problems; otherwise scientists should decide what to investigate. (2)
C. All parties should have an equal say. Government agencies and scientists together should decide what needs to be studied, even though scientists are usually informed about society's needs. (4)
D. Scientists should mostly decide what to investigate, because they know what needs to be studied. Community or government agencies usually know little about science; their advice however, might sometimes be helpful. (2)
E. Scientists should mostly decide because they know best: which areas are ready for a break-through, which areas have the experts available, which areas have the available technology, and which areas have the greatest chance of helping society. (3)
F. Scientists should decide what to investigate, because they alone know what needs to be studied. Governments often put their own interests ahead of society's needs. (1)
G. Scientists should be free to decide what to investigate, because they must be interested in their work in order to be creative and successful. (1)
H. I don't understand. (0)
I. I don't know enough about this subject to make a choice. (0)
J. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis
Appendix 2.10 Item Number 10

Australian scientists should be held responsible for the harm that might result from their discoveries. Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

A. Scientists should be held responsible because it's part of a scientist’s job to ensure that no harm comes from a discovery. Science should cause no harm. (2)
B. Scientists should be held responsible because, if a discovery can be used for both good and bad purposes, the scientists must promote the good use and stop the bad use. (2)
C. Scientists should be held responsible because they must be aware of the effects of their experiments ahead of time. Science should cause more good than harm. (4)
D. The responsibility should be shared about equally between the scientists and society. (3)

Scientists should NOT be held responsible:

E. because it's the people who use the discoveries who are responsible. Scientists may be concerned, but they have no control over how others use their discoveries. (2)
F. because the results of scientific work can't be foreseen (we can't predict if the results will be harmful or not). It's a chance we have to take. (2)
G. because otherwise scientists would quit doing research and science would not progress. (1)
H. because once a discovery is made, others should check its effects. The scientist’s job is only to make the discoveries. Science and moral questions are separate. (1)
I. I don't understand. (0)
J. I don't know enough about this subject to make a choice. (0)
K. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.11 Item Number 11

The best scientists are those who follow the steps of the scientific method. Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

A. The scientific method ensures valid, clear, logical and accurate results. Thus, most scientists will follow the steps of the scientific method. (2)
B. The scientific method should work well for most scientists; based on what we learned in school. (1)
C. The scientific method is useful in many instances. but it does not ensure results. Thus, the best scientists will also use originality and creativity. (4)
D. The best scientists are those who use any method that might get favourable results (including the method of imagination and creativity). (4)
E. Many scientific discoveries were made by accident, and not by sticking to the scientific method. (3)
F. I don't understand. (0)
G. I don't know enough about this subject to make a choice. (0)
H. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.12 Item Number 12

Within Australia there are groups of people who feel strongly in favour of or strongly against some research field. Science and technology projects are influenced by these special interest groups (such as environmentalists, religious organisations, and animal rights people). Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

Special interest groups do have an influence:

A. because they have the power to stop some research projects and that field of science suffers. (2)
B. because they have the power to tell scientists which projects are important to do or not to do. (2)
C. because they influence public opinion and therefore the scientists. (3)
D. because they influence government policy and governments decide whether to fund a research project or not. (4)
E. because some special interest groups give money for certain research projects. (2)
F. Special interest groups try to have an influence but they don't always succeed because scientists and technologists have the final say. (2)
Special interest groups do NOT have an influence:
G. because the government decides the direction that research will take. (1)
H. because science and government decide what projects are important and they do them no matter what special interest groups say. (1)
I. I don't understand. (0)
J. I don't know enough about this subject to make a choice. (0)
K. None of these choices fits my basic viewpoint. (0)  

Appendix 2.13 Item Number 13

Today in Australia, there are more male scientists than female scientists. The MAIN reason for this is: Your position, basically: (Please read from A to K, and then choose ONE only from this page.)
A. males are stronger, faster, brighter, and better at concentrating on their studies. (1)
B. males seem to have more scientific abilities than females, who may excel in other fields. (1)
C. males are just more interested in science than females. (2)
D. the traditional stereotype held by society has been that men are smarter and dominant, while women are weaker and less logical. This prejudice has caused more men to become scientists, even though females are just as capable in science as males. (3)
E. the schools have not done enough to encourage females to take science courses. Females are just as capable in science as males. (2)
F. until recently, science was thought to be a man's vocation. (Women didn't fit television's stereotype image of scientist.) In addition, most women were expected to work in the home or take on traditional jobs. (Thus men have had more encouragement to become scientists.) But today this is changing. Science is becoming a vocation for women, and women are expected to work in science more and more. (4)
G. women have been discouraged, or not allowed, to enter the scientific field. Women are just as interested and just as capable as men; but the established scientists (who are male) tend to discourage or intimidate potential female scientists. (3)
H. There are NO reasons for having more male scientists than female scientists. Both sexes are equally capable of being good scientists, and today the opportunities are equal. (2)
I. I don't understand. (0)
J. I don't know enough about this subject to make a choice. (0)
K. None of these choices fits my basic viewpoint. (0)  

Appendix 2.14 Item Number 14

Science and technology can NOT help people make legal decisions; for example, deciding if a person is guilty or not guilty in a court of law. Your position, basically: (Please read from A to G, and then choose ONE only from this page.)
Science and technology can NOT help:
A. because they have nothing to do with legal decisions, since legal decisions are based on moral values and beliefs. (1)
B. because it's wrong to base legal decisions on technology such as the lie detector. (2)
Science and technology CAN help in a number of cases:
C. by developing ways to gather evidence and by testifying about the physical facts of a case. (4)
D. by studying human behaviour and explaining the human circumstances of a case. (3)
E. I don't understand. (0)
F. I don't know enough about this subject to make a choice. (0)
G. None of these choices fits my basic viewpoint. (0)  

Appendix 2.15 Item Number 15

Scientific research would be better off in Australia if the research were more closely controlled by corporations (for example, companies in high-technology, communications, pharmaceuticals, forestry, mining, manufacturing). Your position, basically: (Please read from A to I, and then choose ONE only from this page.)
Corporations should mainly control science:
A. because closer control by corporations would make science more useful and cause discoveries to be made more quickly through faster communication, better funding, and more competition. (2)
B. in order to improve the cooperation between science and technology, and thus solve problems together. (2)
C. but the public or government agencies should have a say in what science tries to achieve. (2)

Corporations should not control science:
D. because if corporations did, scientific discoveries would be restricted to those discoveries that benefit the corporation (for example, making a profit). Important scientific discoveries that benefit the public are made by unrestricted pure science. (4)
E. because if corporations did, corporations would obstruct scientists from investigating important problems which the companies wanted kept quiet; for example, pollution by the corporation. (3)
F. Science cannot be controlled by corporations. No one, not even the scientist, can control what science will discover. (1)

Appendix 2.16 Item Number 16
Scientists and engineers should be the ones to decide what types of energy Australia will use in the future (for example, nuclear, hydro, solar, or coal burning) because scientists and engineers are the people who know the facts best.
Your position, basically: (Please read from A to J, and then choose ONE only from this page.)
Scientists and engineers should decide:
A. because they have the training and facts which give them a better understanding of the issue. (1)
B. because they have the knowledge and can make better decisions than government bureaucrats or private companies, both of whom have vested interests. (1)
C. because they have the training and facts which give them a better understanding; BUT the public should be involved, either informed or consulted. (2)
D. The decision should be made equally; viewpoints of scientists and engineers, other specialists, and the informed public should all be considered in decisions which affect our society. (4)
E. The government should decide because the issue is basically a political one; BUT scientists and engineers should give advice. (3)
F. The public should decide because the decision affects everyone; BUT scientists and engineers should give advice. (2)
G. The public should decide because the public serves as a check on the scientists and engineers. Scientists and engineers have idealistic and narrow views on the issue and thus pay little attention to consequences. (2)
H. I don't understand. (0)
I. I don't know enough about this subject to make a choice. (0)
J. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.17 Item Number 17
We always have to make trade-offs (compromises) between the positive and negative effects of science and technology.
Your position, basically: (Please read from A to K, and then choose ONE only from this page.)
There are always trade-offs between benefits and negative effects:
A. because every new development has at least one negative result. If we didn't put up with the negative results, we would not progress to enjoy the benefits. (3)
B. because scientists cannot predict the long-term effects of new developments, in spite of careful planning and testing. We have to take the chance. (3)
C. because things that benefit some people will be negative for someone else. This depends on a person's viewpoint. (4)

Scores for responses are shown in parenthesis
D. because you can’t get positive results without first trying a new idea and then working out its negative effects. (2)
E. but the trade-offs make no sense. (For example: Why invent labour saving devices which cause more unemployment? or Why defend a country with nuclear weapons which threaten life on earth?) (2)

There are NOT always trade-offs between benefits and negative effects:
F. because some new developments benefit us without producing negative effects. (2)
G. because negative effects can be minimised through careful planning and testing. (4)
H. because negative effects can be eliminated through careful planning and testing. Otherwise, a new development is not used. (1)
I. I don’t understand. (0)
J. I don’t know enough about this subject to make a choice. (0)
K. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.18 Item Number 18

The best scientists are always very open-minded, logical, unbiased and objective in their work. These personal characteristics are needed for doing the best science.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)
A. The best scientists display these characteristics otherwise science will suffer. (2)
B. The best scientists display these characteristics because the more of these characteristics you have, the better you’ll do at science. (2)
C. These characteristics are not enough. The best scientists also need other personal traits such as imagination, intelligence and honesty. (4)

The best scientists do NOT necessarily display these personal characteristics:
D. because the best scientists sometimes become so deeply involved, interested or trained in their field, that they can be closed-minded, biased, subjective and not always logical in their work. (2)
E. because it depends on the individual scientist. Some are always open-minded, objective, etc. in their work; while others can be come closed-minded, subjective, etc. in their work. (3)
F. The best scientists do NOT display these personal characteristics any more than the average scientist. These characteristics are NOT necessary for doing good science. (1)
G. I don’t understand. (0)
H. I don’t know enough about this subject to make a choice. (0)
I. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.19 Item Number 19

Politics in Australia affects Australian scientists, because scientists are very much part of Australian society (that is, scientists are not isolated from society).

Your position, basically: (Please read from A to J and then choose ONE only from this page.)
Scientists ARE affected by Australian politics:
A. because funding for science comes mainly from governments which control the way the money is spent. Scientists sometimes have to lobby for funds. (4)
B. because governments not only give money for research, they set policy regarding new developments. This policy directly affects the type of projects scientists will work on. (4)
C. because scientists are a part of society and are affected like everyone else. (3)
D. because scientists try to help society and thus they are closely tied to society. (2)

Scientists are NOT affected by Australian politics:
E. because the nature of a scientist’s world prevents the scientist from becoming involved politically. (1)
F. because scientists are isolated from society; their work receives no public media attention unless they make a spectacular discovery. (1)
G. because Australia is a free country, and so scientists can work quite freely. (1)
H. I do not understand . (0)
I. I don’t know enough about this subject to make a choice. (0)
J. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis
Appendix 2.20 Item Number 20

When doing science or technology, a good female scientist would carry out the job basically in the same way as a good male scientist.

Your position, basically: (Please read from A to L, and then choose ONE only from this page.)

There is NO difference between female and male scientists in the way they do science:

A. because all good scientists carry out the job the same way. (2)
B. because female and male scientists experience the same training. (3)
C. because overall, women and men are equally intelligent. (2)
D. because women and men are the same in terms of what is needed to be a good scientist. (3)
E. because everyone is equal, no matter what the job. (1)
F. because any differences in the way scientists do science are due to differences between individuals. Such differences have nothing to do with being male or female. (4)
G. Women would do science somewhat differently because, by nature or by upbringing, females have different viewpoints, perspectives, imagination, or characteristics (such as patience). (3)
H. Men would do science somewhat differently because men do science better. (1)
I. Women would likely do science somewhat better than men, because women must work harder in order to compete in a male dominated field such as science. (2)
J. I don't understand. (0)
K. I don't know enough about this subject to make a choice. (0)
L. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.21 Item Number 21

The success of science and technology in Australia depends on how much support the public gives to scientists, engineers and technicians. This support depends on high school students, the future public, learning how science and technology are used in Australia.

Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

Yes, the more students learn about science and technology:

A. the better they will keep the country running. High school students are the future. (2)
B. the more students will become scientists, engineers and technicians, and so Australia will prosper. (2)
C. the more informed the future public will be. They will be able to form better opinions and make better contributions to how science and technology are used. (4)
D. the more the public will see that science and technology are important. The public will better understand the views of experts and will provide the needed support for science and technology. (3)
E. No, support does not depend on students learning more about science and technology. Some high school students are not interested in science subjects. (1)
F. I don't understand. (0)
G. I don't know enough about this subject to make a choice. (0)
H. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.22 Item Number 22

More technology will improve the standard of living for Australians.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. Yes, because technology has always improved the standard of living, and there is no reason for it to stop now. (1)
B. Yes, because the more we know, the better we can solve our problems and take care of ourselves. (2)
C. Yes, because technology creates jobs and prosperity. Technology helps life become easier, more efficient and more fun. (2)
D. Yes, but only for those who can afford to use it. More technology will cut jobs and cause more people to fall below the poverty line. (3)
E. Yes and no. More technology would make life easier, healthier and more efficient. BUT more technology would cause more pollution, unemployment and other problems. The standard of living may improve, but the quality of life may not. (4)

F. No. We are irresponsible with the technology we have now; for example, our production of weapons and using up our natural resources. (3)

G. I don't understand (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0)  

Scores for responses are shown in parenthesis

Appendix 2.23 Item Number 23

Few scientists and technologists would choose to work on military research and development.  
Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

Few would do research and development for the military:
A. because many scientists and technologists would rather work in other areas which benefit human life and the environment. (2)
B. because many scientists and technologists would not sacrifice their morals and contribute to the violence of war. (2)
C. It depends on the person's values and research interests. Some scientists would find the military projects interesting and rewarding; other scientists would rather not work on projects related to war. (4)

A number of scientists and technologists choose to work for the military:
D. because most of the research money is in arms technology and military related research. The military offers large budgets, excellent equipment and more recognition for scientists. (3)
E. because they know that our country's defence is important. We need more scientists in military research and development. (1)
F. I don't understand. (0)
G. I don't know enough about this subject to make a choice. (0)
H. None of these choices fits my basic viewpoint. (0)  

Scores for responses are shown in parenthesis

Appendix 2.24 Item Number 24

We have to be concerned about pollution problems which are unsolvable today. Science and technology cannot necessarily fix these problems in the future.  
Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

Science and technology can NOT fix such problems:
A. because science and technology are the reason that we have pollution problems in the first place. More science and technology will bring more pollution problems. (1)
B. because pollution problems are so bad today they are already beyond the ability of science and technology to fix them. (1)
C. because pollution problems are becoming so bad that they may soon be beyond the ability of science and technology to fix them. (1)
D. No one can predict what science and technology will be able to fix in the future. (3)
E. Science and technology alone cannot fix pollution problems. It is everyone's responsibility. The public must insist that fixing these problems is a top priority. (4)
F. Science and technology can fix such problems, because the success at solving problems in the past means science and technology will be successful in the future at fixing pollution problems. (2)
G. I don't understand. (0)
H. I don't know enough about this subject to make a choice. (0)
I. None of these choices fits my basic viewpoint. (0)  

Scores for responses are shown in parenthesis

Appendix 2.25 Item Number 25

Science would advance more efficiently in Australia if it were more clearly controlled by the government.  
Your position, basically: (Please read from A to H, and then choose ONE only from this page.)
Appendix 2.26 Item Number 26

Science and technology influence our everyday thinking because science and technology give us new words and ideas.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. Yes, because the more you learn about science and technology, the more your vocabulary increases, and thus the more information you can apply to everyday problems. (2)
B. Yes, because we use the products of science and technology (for example, computers, microwaves, health care). New products add new words to our vocabulary and change the way we think about everyday things. (2)
C. Science and technology influence our everyday thinking BUT the influence is mostly from new ideas, inventions and techniques which broaden our thinking. (2)
D. but because almost everything we do, and everything around us, has in some way been researched by science and technology. (4)
E. but because science and technology have changed the way we live. (3)
F. No, because our everyday thinking is mostly influenced by non-scientific things. Science and technology influence only a few of our ideas. (1)
G. I don't understand. (0)
H. I don't know enough about this subject to make a choice. (0)
I. None of these choices fits my basic viewpoint. (0)  Scores for responses are shown in parenthesis

Appendix 2.27 Item Number 27

Most Australian scientists are motivated to work hard. The MAIN reason behind their personal motivation for doing science is:

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

A. earning recognition, otherwise their work would not be accepted. (3)
B. earning money, because society pressures scientists to strive after financial rewards. (2)
C. acquiring a bit of fame, fortune and power, because scientists are like anyone else. (2)
D. satisfying their curiosity about the natural world, because they like to learn more all the time and solve mysteries of the physical and biological universe. (3)
E. solving curious problems for personal knowledge, AND discovering new ideas or inventing new things that benefit society (for example, medical cures, answers to pollution, etc.). Together these represent the main personal motivation of most scientists. (4)
F. unselfishly inventing and discovering new things for technology. (1)
G. discovering new ideas or inventing new things that benefit society (for example, medical cures, answers to pollution, etc.). (1)
H. It’s not possible to generalise because the main personal motivation of scientists varies from scientist to scientist. (4)
I. I don't understand. (0)
APPENDIX 3: THE FIT DATA FOR THE ITEMS

Appendix 3.1 Item estimates (thresholds) in input order:
students on Science Scale (N = 1278; L = 10)

<table>
<thead>
<tr>
<th>Item Name</th>
<th>Score</th>
<th>Max Score</th>
<th>Thresholds 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Inft t</th>
<th>Oult t</th>
<th>MnSq</th>
<th>Mn Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>3000</td>
<td>5112</td>
<td>-1.06</td>
<td>-.78</td>
<td>.46</td>
<td>2.02</td>
<td>1.09</td>
<td>1.09</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.16</td>
<td>.13</td>
<td>.12</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>2995</td>
<td>5112</td>
<td>-0.89</td>
<td>-.70</td>
<td>.68</td>
<td>1.05</td>
<td>1.10</td>
<td>1.09</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.12</td>
<td>.09</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>3540</td>
<td>5112</td>
<td>-0.91</td>
<td>-.63</td>
<td>.05</td>
<td>0.73</td>
<td>0.96</td>
<td>0.95</td>
<td>-1.0</td>
<td>-0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.12</td>
<td>.11</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 10</td>
<td>2803</td>
<td>5112</td>
<td>-1.28</td>
<td>-.70</td>
<td>.90</td>
<td>1.40</td>
<td>1.10</td>
<td>1.10</td>
<td>2.3</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.14</td>
<td>.13</td>
<td>.11</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 14</td>
<td>3576</td>
<td>5112</td>
<td>-0.50</td>
<td>-.13</td>
<td>.06</td>
<td>0.26</td>
<td>0.97</td>
<td>0.96</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.09</td>
<td>.11</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 16</td>
<td>2966</td>
<td>5112</td>
<td>-1.31</td>
<td>-.15</td>
<td>.58</td>
<td>0.73</td>
<td>0.94</td>
<td>0.95</td>
<td>-1.7</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.11</td>
<td>.10</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 17</td>
<td>3239</td>
<td>5112</td>
<td>-0.89</td>
<td>-.39</td>
<td>.17</td>
<td>1.03</td>
<td>1.05</td>
<td>1.07</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 22</td>
<td>3397</td>
<td>5112</td>
<td>-1.13</td>
<td>-.32</td>
<td>.21</td>
<td>0.59</td>
<td>0.92</td>
<td>0.90</td>
<td>-2.4</td>
<td>-2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.10</td>
<td>.09</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 24</td>
<td>3277</td>
<td>5112</td>
<td>-0.91</td>
<td>-.23</td>
<td>.17</td>
<td>-.75</td>
<td>-.93</td>
<td>-.92</td>
<td>-2.1</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.12</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 26</td>
<td>2655</td>
<td>5112</td>
<td>-0.75</td>
<td>-.46</td>
<td>.91</td>
<td>1.34</td>
<td>1.01</td>
<td>1.01</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.12</td>
<td>.12</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean 0.00 1.01 1.01 0.3 0.2 SD 0.14 0.07 0.08 1.9 1.5

Appendix 3.2 Item estimates (thresholds) in input order:
students on Scientist Scale (N = 1278; L = 8)

<table>
<thead>
<tr>
<th>Item number</th>
<th>Score</th>
<th>Max Score</th>
<th>Threshold 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Inft t</th>
<th>Oult t</th>
<th>MnSq</th>
<th>Mn Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 4</td>
<td>3229</td>
<td>5112</td>
<td>-1.25</td>
<td>-.60</td>
<td>.53</td>
<td>.91</td>
<td>0.99</td>
<td>0.98</td>
<td>-0.2</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.16</td>
<td>.13</td>
<td>.10</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>3337</td>
<td>5112</td>
<td>-0.63</td>
<td>0.01</td>
<td>.15</td>
<td>.58</td>
<td>1.09</td>
<td>1.14</td>
<td>2.7</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.09</td>
<td>.11</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td>2796</td>
<td>5112</td>
<td>-0.67</td>
<td>-.18</td>
<td>.59</td>
<td>1.34</td>
<td>1.07</td>
<td>1.08</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.10</td>
<td>.09</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 11</td>
<td>3510</td>
<td>5112</td>
<td>-0.39</td>
<td>-.16</td>
<td>.08</td>
<td>.46</td>
<td>0.99</td>
<td>0.98</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 13</td>
<td>3142</td>
<td>5112</td>
<td>-1.42</td>
<td>-.70</td>
<td>.52</td>
<td>1.28</td>
<td>0.95</td>
<td>0.95</td>
<td>-1.2</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.17</td>
<td>.14</td>
<td>.09</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 18</td>
<td>3058</td>
<td>5112</td>
<td>-0.69</td>
<td>-.30</td>
<td>.43</td>
<td>1.08</td>
<td>0.99</td>
<td>1.00</td>
<td>-0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.10</td>
<td>.09</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 20</td>
<td>3189</td>
<td>5112</td>
<td>-1.22</td>
<td>-.55</td>
<td>.43</td>
<td>1.14</td>
<td>0.93</td>
<td>0.93</td>
<td>-2.0</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.16</td>
<td>.13</td>
<td>.10</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 27</td>
<td>3727</td>
<td>5112</td>
<td>-0.87</td>
<td>-.41</td>
<td>.92</td>
<td>0.52</td>
<td>1.00</td>
<td>0.98</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>.11</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean 0.00 1.00 1.00 0.1 0.0 SD 0.15 0.05 0.07 1.5 1.3
### APPENDIX 4: MAP OF RESPONDENTS AND ITEM ESTIMATES (THRESHOLDS)

Appendix 4.1 Map of respondents and item estimates (thresholds):
all students on Science Scale (N = 1278; L = 10)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
</tr>
<tr>
<td>xx</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>xxx</td>
<td>10.4</td>
</tr>
<tr>
<td>xxxx</td>
<td>26.4</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>xxxxxxx</td>
<td>2.4</td>
</tr>
<tr>
<td>xxxxxxxxx</td>
<td>17.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>10.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>26.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>2.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>7.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>16.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>24.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>16.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>22.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>1.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>14.4</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>17.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>22.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>24.3</td>
</tr>
<tr>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>7.3</td>
</tr>
<tr>
<td>xxxxxxxxxxxxxx</td>
<td>14.3</td>
</tr>
<tr>
<td>xxxxxxx</td>
<td>14.2</td>
</tr>
<tr>
<td>xxxxxxx</td>
<td>16.2</td>
</tr>
<tr>
<td>xx</td>
<td>22.2</td>
</tr>
<tr>
<td>xxx</td>
<td>24.2</td>
</tr>
<tr>
<td>x</td>
<td>17.2</td>
</tr>
<tr>
<td>xx</td>
<td>14.1</td>
</tr>
<tr>
<td>x</td>
<td>26.2</td>
</tr>
<tr>
<td>xx</td>
<td>2.2</td>
</tr>
<tr>
<td>x</td>
<td>10.2</td>
</tr>
<tr>
<td>x</td>
<td>26.1</td>
</tr>
<tr>
<td>-1.0</td>
<td>x</td>
</tr>
<tr>
<td>x</td>
<td>1.2</td>
</tr>
<tr>
<td>x</td>
<td>2.1</td>
</tr>
<tr>
<td>x</td>
<td>7.1</td>
</tr>
<tr>
<td>x</td>
<td>17.1</td>
</tr>
<tr>
<td>x</td>
<td>24.1</td>
</tr>
</tbody>
</table>

| -2.0 |   |
| x |   |

Each X represents 8 students
Appendix 4.2  Map of respondents and item estimates (thresholds): all students on Scientists Scale (N = 1426; L = 8)

Each X represents 11 students