WAT04867

Student Change in Understanding of Statistical Variation after Instruction and after Two Years: An Application of Rasch Analysis

Jane M. Watson and Ben A. Kelly, University of Tasmania
John F. Izard, RMIT University

Abstract

Data collected from students involved in a project examining change in understanding of statistical variation in relation to the chance and data curriculum after instruction and after two years, are the basis for the analysis reported in this study. Comparisons are made, using partial credit Rasch analysis, between successive grades (3, 5, 7, and 9), within students after instruction, within students after two years, and between students in the four grades after two years depending on whether they were involved in the instructional intervention or not. Results show varying magnitudes of differences among grades, differing improvements after instruction, but little difference between Intervention and Non-Intervention groups after a two year period.

Introduction

The data used in this study were collected as part of a three-year research project considering students' understanding of variation as part of the chance and data school curriculum. The project itself arose as a second phase of research following the introduction of chance and data as a formal part of the mathematics curriculum in the early 1990s (Australian Education Council [AEC], 1991). This move had followed a similar one in the United States, initiated by the National Council of Teachers of Mathematics (NCTM, 1989). These curriculum documents suggested organisers for student learning such as chance outcomes, measurement of middles and spread, the production of visual displays of data, the collection of samples, and the drawing of inferences. Although all of these topics depend upon variation of some sort for their interest, explicit focus on variation is missing from the documents. Following initial research on topics such as average (e.g., Mokros & Russell, 1995), probability (e.g., Green, 1986, 1991; Fischbein & Gazit, 1984), graphing (Curcio, 1987), and inference (Watson, Collis, Callingham, & Moritz, 1995), calls from prominent researchers in the field (e.g., Green, 1993; Shaughnessy, 1997) to consider variation more specifically were heeded in several research projects around the world (e.g., Cobb, 1999; Shaughnessy, Canada, & Ciancetta, 2003), including the Australian one described here.

The premise of the research follows the work of Moore (1990) in his emphasis on variation as the foundation of all of statistics and probability. The methods that students learn reflect ways of finding, handling, measuring, showing, controlling, and/or explaining variation in uncertain situations where data are collected. Because variation occurs for data in a context, it is difficult to consider variation in an abstract way on its own. One of the goals of the project from which the data used in this study were collected was to provide learning experiences for students based on contexts from the chance and data curriculum as articulated in Australia (AEC, 1991) and Tasmania (Department of Education and the Arts, 1992) that would also emphasise the importance of variation. This was done in the primary grades (3 and 5) by the project providing a teacher to engage with students in 10 sessions in their classrooms. Activities included work with small boxes of coloured sweets, with spinners and dice,

with sampling, and with measurement, finishing with students planning their own investigations. These activities are described in detail in Watson and Kelly (2002a). In the secondary grades (7 and 9) a series of six short units were provided to the mathematics teachers in the participating schools. After discussion with the researchers the teachers chose from work based on spinners, dice, sampling, measurement data for considering association and comparing two groups, and the number of chocolate chips in chocolate chip cookies.

Before and after the instructional intervention pre and post tests were administered to students and paired *t*-tests showed that both overall and on subscales associated with basic chance and data, variation in chance, variation in data and graphing, and variation in sampling, students in each grade improved after the intervention (Watson & Kelly, 2002a, 2002b, 2002c). Other students were also involved in the project, in schools matched with the intervention schools for socio-economic status. These students completed the pre test but did not experience any lessons organised or suggested by members of the project team. They did not complete a post test. Two years later all students who could be traced were given a longitudinal follow-up test, identical to the pre and post tests. Preliminary comparisons using conventional measurement techniques based on *t*-tests are reported in Watson and Kelly (in press). They found mixed results after two years, with Grades 3 and 7 in both Intervention and Non-Intervention schools showing significant positive change, Grade 5 in both types of schools showing no change, and Grade 9 showing improvement for Non-Intervention schools but not for Intervention schools.

Given the earlier work of Watson, Kelly, Callingham, and Shaughnessy (2003), to describe the understanding of statistical variation on a hierarchical scale based on Rasch analysis using the complete data set at the beginning of the project, it is of interest to follow this with Rasch analysis of data for the post test for the intervention students and for the longitudinal test for all students who could be contacted.

The research questions hence focus on the observed difference in understanding across the grades, change in understanding observed for the students in the instructional Intervention on the post test, and change in understanding after two years for both Intervention and Non-Intervention students. It is then possible to compare the two groups of students after two years.

Methodology

Sample

The original sample consisted of 738 students from ten state government schools in the Australian state of Tasmania. Students were in Grades 3, 5, 7, and 9 at the beginning of the study in 2000. Five of the ten schools took part in the instructional Intervention described in the Introduction. Two of the schools were high schools with one primary school being a feeder school to one high school and two primary schools being feeder schools to the other high school. This arrangement made it more likely that students could be followed from Grade 5 to Grade 7 two years later. All Grade 9 students either changed schools to a senior secondary college or dropped out of school in the two years before the longitudinal survey. Similarly for the other five Non-Intervention schools in the study, three were primary and two were high schools, with the primaries being feeders to the high schools in the study. The numbers of students at each group at each stage are given in Table 1. All students in

subsequent groups had completed the previous survey or surveys. In the primary schools all students in Grades 3 and 5 participated in the study. In Grades 7 and 9 all students in selected classes participated in the study. The range of performance observed indicated that no students were withdrawn by their teachers unknown to the researchers.

Table 1
Number of Students in Each Grade at the Time of Each Survey with Complete Data

	Intervention				Non-Intervention				
	G3/5[1]	G5/7[1]	G7/9[1]	G9/11[1]	G3/5[1]	G5/7[1]	G7/9[1]	G9/11[1]	TOTAL
Pre	85	90	105	109	91	93	81	84	738
Pre-Post	72	82	93	90					337
Pre-Post-Long	47	58	67	28	67	44	68	31	410

[1]Grade in the longitudinal follow-up

Instrument

The survey instrument consisted of 50 items, however, not all were presented to all grades. Grade 3 students received 25 items, Grade 5 students received 29 items, Grade 7 students received 45 items, and Grade 9 students received 46 items. Additional items were included to increase overall complexity with grade. All items are presented in the Appendix to Watson et al. (2003). The items were devised to address four aspects of chance and data: basic chance and data (BCD, 14 items), variation in chance (CV, 11 items), variation in data and graphing (DV, 14 items), and variation in sampling (SV, 12 items), with one item contributing to two subscales.

Procedure

The surveys were administered by the first two authors with the classroom teachers, all offering help where required to read items, particularly in Grades 3 and 5. The surveys were not intended as reading tests but no indication of appropriate answers was provided if help with reading was given. Approximately 45 minutes was allocated for completing the surveys. At each administration of the instrument students were given the survey form devised for that grade. Hence in 2002, students originally in Grade 5 in 2000, answered more questions, as found in the Grade 7 form of the survey. For comparisons carried out in this study, analyses are based on the scaled Rasch estimates of achievement for the groups being compared.

Coding

The coding of items is described in detail in Watson et al. (2003) and was based on the structure of the observed responses (SOLO) (Biggs & Collis, 1982, 1991), the statistical appropriateness of responses, and coding schemes from earlier studies that had included some of the items. Although some items were related to the same stem, they were coded independently and they were developed in such a way that a correct response to a particular item was not a prerequisite to a correct response to the next or another succeeding item. This coding system hence met the criteria for the application of the Rasch model. Thirteen items were coded on a 0-1 (incorrect-correct) basis and the rest followed a partial credit model, with successively higher codes assigned to more appropriate and often more complex responses.

Analysis

The data were analysed using Rasch (1980) measurement techniques, which allowed both students' performances and item difficulties to be measured using the same metric, and placed on the same scale. The Quest computer program (Adams & Khoo, 1996) was used to apply the Partial Credit Model (Masters, 1982) and obtain a variable map showing the placement of students and items along the scale. This procedure was carried out for 2000 pre test data on the items in common across Grades 3, 5, 7 and 9 to provide anchor values for those items. Then items in common to Grades 3 and 5 not already anchored were added to the anchor group in a subsequent analysis. Items tackled by Grade 5 not already anchored were then added using the anchored items. This was followed by an analysis of combined Grade 7 and 9 data that anchored all items in common to Grade 7 and 9 not already anchored. Then the final item attempted by Grade 9 was anchored in a subsequent analysis.

Several statistics, produced by the Quest program, are used to evaluate the fit of the data to the Partial Credit Rasch Model. The first of these is the Infit Mean Square (IMSQ), a measure of the extent to which the fit of the items (item IMSQ) or persons (case IMSQ) deviates from the expected value of 1.00. Acceptable values lie between 0.77 and 1.3 (Adams & Khoo, 1996). There were five separate analyses in the calibration of the items in the overall scale. The values for items (mean item IMSQ), and persons (mean case IMSQ) are shown in Table 2. In all instances the values in this study were acceptable. The Separation Reliability is a measure of how well the items (R_I) or persons (R_P) behave consistently. These statistics may be interpreted as a reliability statistic, and have an ideal value of 1. The values for both items (R_I) and persons (R_P) were high, indicating that the behaviours of both items and persons were consistent. (Details for items are shown in Appendix A.)

Table 2
Fit and Separation Reliability Values for the 2000 Data by Analysis

Analysis Details	Number of Persons	Number of Items	Mean Item IMSQ	Mean Person IMSQ	R_I (rounded)	R_P (rounded)
Items 1 to 21 (Run No 1) [Items in common]	738	21	1.00	1.05	0.98	0.85
Items 1 to 25 (Run No B1) [Grades 3 and 5]	359	25	0.90	0.95	0.99	0.86
Items 1 to 29 (Run No B2) [Grade 5]	183	29	0.86	0.92	0.98	0.81
Items 1 to 21 and 26 to 49 (Run No C1) [Grades 7 and 9]	379	45	1.05	1.10	0.98	0.92
Items 1 to 21 and 26 to 50 (Run No C2) [Grade 9]	193	46	1.08	1.10	1.00	0.91

The anchor values were used in similar sets of analyses to establish a scaled score for each student in the post test group. Then the procedure was repeated in a

similar set of analyses to establish a scaled score for each student in the longitudinal group.

The scaled achievement scores for each student on each test could then be used to determine the initial differences among the grades and the changes over the time between testings. The effect sizes for these differences were determined using Cohen's (1969) methodology and reported with descriptors devised by Cohen (1969) and Izard (2004). These are shown in Table 3.

Table 3
Descriptors for Magnitudes of Effect Sizes (after Cohen, 1969, p.23) and Assigned Ranges

Effect Size Magnitude	Cohen's Descriptor and Cohen's Example	Assigned Range
< 0.2	Very small*	0.00 to 0.14
0.2	Small difference between the heights of 15 year old and 16 year old girls in the US	0.15 to 0.44
0.5	Medium ('large enough to be visible to the naked eye') difference between the heights of 14 year old and 18 year old girls	0.45 to 0.74
0.8	Large ('grossly perceptible and therefore large') difference between the heights of 13 year old and 18 year old girls or the difference in IQ between holders of the Ph.D. degree and 'typical college freshmen'	0.75 or more

* Note that "very small" is a descriptor devised by Izard (2004) for magnitudes less than "small".

Results

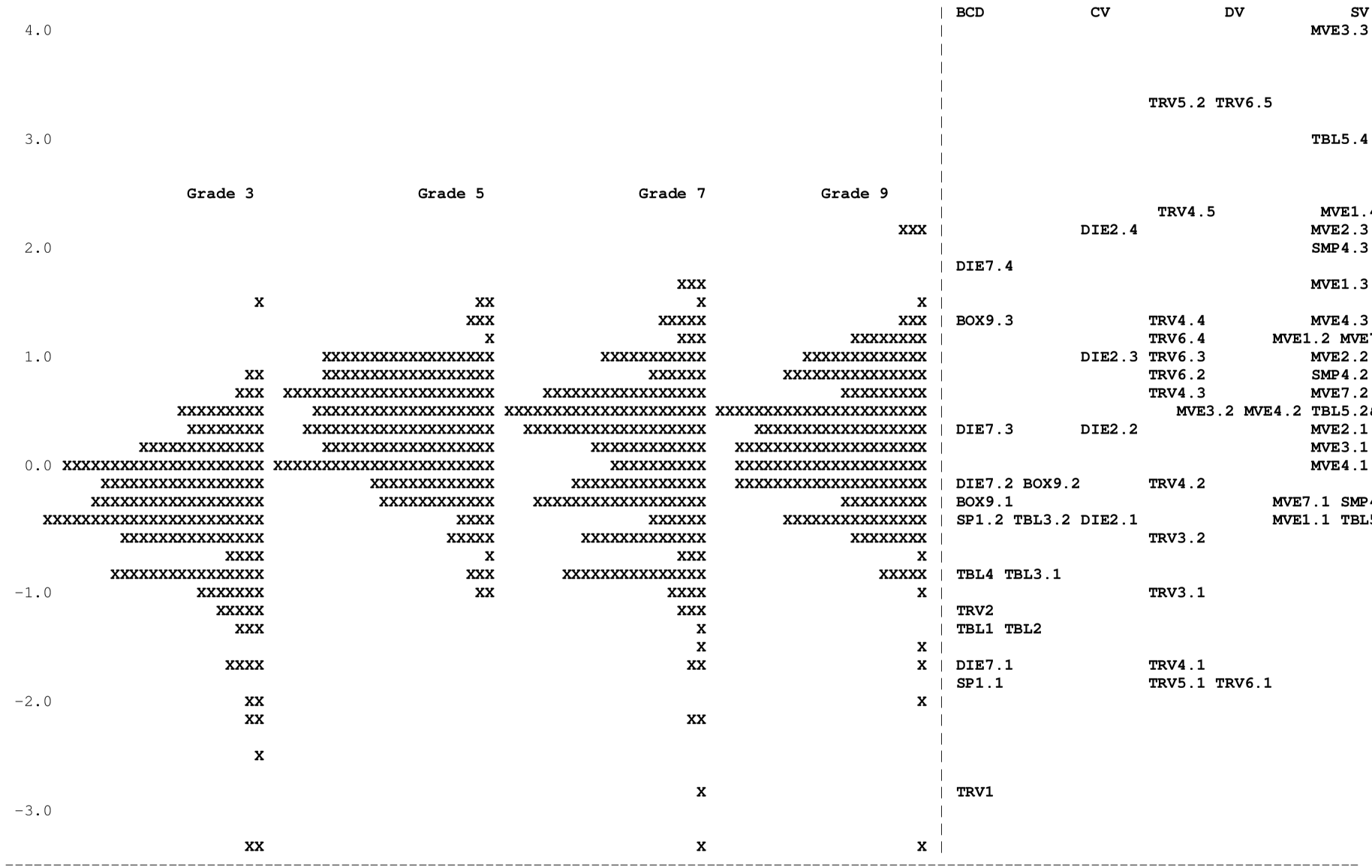
The results are presented in three parts. The first part considers a brief description of the common items answered by all students and change observed across the four grades at the beginning of the study. The second is a consideration of change after the post and longitudinal tests for the Intervention group and after the longitudinal test for the Non-Intervention group. Third, results are compared for the two groups in terms of the longitudinal change.

Part 1 – Across Grade Comparisons

Figure 1 shows the variable map of performance across Grades 3, 5, 7, and 9 for the 21 common items answered by all students. The items on the right have been placed in columns in relation to the subgroups of items described in the Instrument section. The spread of the items suggests that the basic items (BCD) were somewhat easier than the data variation items (DV), and again than the sampling variation items (SV). The easiest items reflected basic table and graph reading and an intuitive appreciation of chance.

As can be seen on the left side of Figure 1, there is a movement upward in the distribution of students' abilities from Grade 3 to Grade 5, but little apparent change with grade after that. The spread of the distributions increases after Grade 5, with the top achieving higher scores each time but still some performance at the bottom tail of the distribution.

Figure 1. Items 1 to 21 (Items in common) 2000 data [Item Estimates (Thresholds) all on all (N = 738 L = 21 Probability Level=0.50)]



Each X represents 1 student

Note: BCD=Basic Chance & Data; CV=Chance Variation; DV=Data Variation; SV=Sampling Variation

Table 4 presents the mean logit scores for each grade (N) and the successive grade (N+2) for the students in the Intervention group. The comparisons are made for two groups of students, those who had complete data for all three surveys in the study and all students who completed the survey at the beginning of the project. The first group was a subset of the second group. Except in the comparisons involving Grade 9, the results were similar regardless of the sample used. As can be seen, confirming the observation in Figure 1, there is a large improvement from Grade 3 to Grade 5 and a regression from Grade 5 to Grade 7. The medium improvement from Grade 7 to Grade 9 found in the smaller group was reduced to a small improvement when the larger group was used. This is likely to reflect the bias in the smaller Grade 9 group representing only students with the ability to continue schooling to Grade 11. The change from Grades 5 to 7 prompted a comparison to be made from Grade 5 to Grade 9, which indicated a very small improvement in the smaller sample and a small negative change in the larger sample. This relationship of differences and sample size is similar to that observed in comparing Grade 7 and 9.

Table 4.
Results for Grade Comparisons within the Intervention Group

Intervention Group	Pre-test (N) vs Pre-test (Year N+2)	
	Long Subset Data	Pre Data Set
Grades 3/5	(*n*=47, *n*=58)	(*n*=85, *n*=90)
Pre test 3 Mean, SD	-0.19, 0.72	-0.27, 0.72
Pre test 5 Mean, SD	0.35, 0.44	0.34, 0.50
Mean Difference	0.54	0.61
Effect Size (SE)	0.92 (0.21)	0.99 (0.16)
Cohen's descriptor	Large	Large
Grades 5/7	(*n*=58, *n*=67)	(*n*=90, *n*=105)
Pre test 5 Mean, SD	0.35, 0.44	0.34, 0.50
Pre test 7 Mean, SD	0.00, 0.75	-0.07, 0.88
Mean Difference	-0.35	-0.41
Effect Size (SE)	-0.56 (0.18)	-0.56 (0.15)
Cohen's descriptor	Medium (negative)	Medium (negative)
Grades 7/9	(*n*=67, *n*=28)	(*n*=105, *n*=109)
Pre test 7 Mean, SD	0.00, 0.75	-0.07, 0.88
Pre test 9 Mean, SD	0.38, 0.62	0.20, 0.74
Mean Difference	0.38	0.27
Effect Size (SE)	0.53 (0.23)	0.33 (0.14)
Cohen's descriptor	Medium	Small
Grades 5/9	(*n*=58, *n*=28)	(*n*=90, *n*=109)
Pre test 5 Mean, SD	0.35, 0.44	0.34, 0.50
Pre test 9 Mean, SD	0.38, 0.62	0.20, 0.74
Mean Difference	0.03	-0.14
Effect Size (SE)	0.06 (0.23)	-0.22 (0.14)
Cohen's descriptor	Very small	Small (negative)

Table 5 shows similar results for each grade (N) and the successive grade (N+2) for the students in the Non-Intervention group. Again results are presented for students who completed both pre and longitudinal surveys, and for students who completed the pre test survey at the beginning of the project. The differences for Grades 3 and 5 are again large for both data sets, as for the Intervention group. The Grade 5 to 7 difference, however, was negative for both samples, either very small or small. For Grades 7 and 9, the results reflect the same trend as for the Intervention

group with a medium change for the small sample and a small (either negative or positive) change for the large sample. For the Non-Intervention group, however, the Grade 9 results were better than the Grade 5 for the smaller subgroup continuing to the end of the study, with an effect judged to be medium. Again for the larger group the difference is small and negative.

Table 5.
Results for Grade Comparison within the Non-Intervention Group

Non-intervention Group	Pre-test (N) vs Pre-test (Year N+2)	
	Long Subset Data	Pre Data Set
Grades 3/5	(*n*=67, *n*=44)	(*n*=91, *n*=93)
Pre test 3 Mean, SD	-0.46, 0.83	-0.48, 0.76
Pre test 5 Mean, SD	0.33, 0.55	0.40, 0.57
Mean Difference	0.79	0.88
Effect Size (SE)	1.09 (0.21)	1.31 (0.16)
Cohen's descriptor	Large	Large
Grades 5/7	(*n*=44, *n*=68)	(*n*=93, *n*=81)
Pre test 5 Mean, SD	0.33, 0.55	0.40, 0.57
Pre test 7 Mean, SD	0.26, 0.62	0.27, 0.60
Mean Difference	-0.07	-0.13
Effect Size (SE)	-0.12 (0.19)	-0.22 (0.15)
Cohen's descriptor	Very small (negative)	Small (negative)
Grades 7/9	(*n*=68, *n*=31)	(*n*=81, *n*=84)
Pre test 7 Mean, SD	0.26, 0.62	0.27, 0.60
Pre test 9 Mean, SD	0.62, 0.50	0.30, 0.61
Mean Difference	0.36	0.03
Effect Size (SE)	0.61 (0.22)	0.05 (0.16)
Cohen's descriptor	Medium	Very small
Grades 5/9	(*n*=44, *n*=31)	(*n*=93, *n*=84)
Pre test 5 Mean, SD	0.33, 0.55	0.40, 0.57
Pre test 9 Mean, SD	0.62, 0.50	0.30, 0.61
Mean Difference	0.29	-0.10
Effect Size (SE)	0.55 (0.24)	-0.17 (0.15)
Cohen's descriptor	Medium	Small (negative)

Part 2 – Change Over Time within the Intervention and Non-Intervention Groups

For the group of 410 students who were in the study for the two years, Tables 6 and 7 show the differences in their average performance over this time. For the Intervention students this includes the performance on the post-test 6 weeks after the completion of the instructional unit, based on two groups, those present for all three surveys (e.g., *n*=47 for Grade 3/5) and those who were present at the post survey (e.g., *n*=72 for Grade 3/5).

Table 6
Results for Comparisons within the Intervention on the Post and Longitudinal Survey

	Pre vs Post		Pre vs Longitudinal
	Long Subset Data	Post Subset Data	Long Subset Data
Grade 3/5[1]	(*n*=47)	(*n*=72)	(*n*=47)
Pre Mean, SD	-0.19, 0.72	-0.19, 0.69	-0.19, 0.72
Post or Long Mean, SD	0.07, 0.82	0.09, 0.74	0.50, 0.50
Mean Diff	0.26	0.28	0.69
Effect Size (SE)	0.33 (0.21)	0.39 (0.17)	1.10 (0.22)
Cohen's descriptor	Small	Small	Large
Grade 5/7[1]	(*n*=58)	(*n*=82)	(*n*=58)
Pre Mean, SD	0.35, 0.44	0.38, 0.46	0.35, 0.44
Post or Long Mean, SD	0.52, 0.50	0.59, 0.52	0.28, 0.62
Mean Diff	0.17	0.21	-0.07
Effect Size (SE)	0.36 (0.19)	0.43 (0.16)	-0.13 (0.19)
Cohen's descriptor	Small	Small	Very small (negative)
Grade 7/9[1]	(*n*=67)	(*n*=93)	(*n*=67)
Pre Mean, SD	0.00, 0.75	0.06, 0.77	0.00, 0.75
Post or Long Mean, SD	0.52, 0.79	0.60, 0.83	0.57, 0.83
Mean Diff	0.52	0.54	0.57
Effect Size (SE)	0.67 (0.18)	0.67 (0.15)	0.72 (0.18)
Cohen's descriptor	Medium	Medium	Medium
Grade 9/11[1]	(*n*=28)	(*n*=90)	(*n*=28)
Pre Mean, SD	0.38, 0.62	0.20, 0.76	0.38, 0.62
Post or Long Mean, SD	0.34, 0.81	0.50, 0.66	0.53, 0.79
Mean Diff	-0.04	0.30	0.15
Effect Size (SE)	-0.05 (0.27)	0.42 (0.15)	0.21 (0.27)
Cohen's descriptor	Very small (negative)	Small	Small

[1]Grade in Longitudinal follow-up

The greatest improvement six weeks after instruction was shown by the Grade 7 students with smaller degrees of improvement shown by the Grade 3 and 5 students; however, it should be noted that the Grade 7 students started with a much lower mean than the Grade 5 students did (see Table 4). For the reduced sample of Grade 9 students who completed the entire study, there was a negligible drop in performance on the post test. This decline was not found, however, for the larger group of students (*n*=90) who completed the post test six weeks after instruction finished. For this group there was a small improvement on the post survey. For the other groups, with a smaller difference in the size of the two groups tested, the same effect sizes were observed. Over the following two years, the Grade 3/5 students showed the greatest improvement, with the Grade 7/9 students retaining their medium level of improvement and the Grade 5/7 students dropping negligibly below their original performance. The Grade 9/11 group improved slightly.

As can be seen in Table 7 for the Non-Intervention group after two years, the change for students' performance was very similar to that for the Intervention group after two years, with the Grade 3/5 students showing a large improvement and the Grade 7/9 students a medium improvement. The smaller Grade 9/11 group showed a medium improvement but again the Grade 5/7 students declined slightly.

Table 7
Results for Comparisons within the Non-Intervention Group

	Pre vs Longitudinal
Grade 3/5[1] (*n*=67)	
Pre Mean, SD	-0.46, 0.83
Longitudinal Mean, SD	0.28, 0.42
Mean [of] Difference [SD]	0.74
Effect Size (SE)	1.12 (0.10)
Cohen's descriptor	Large
Grade 5/7[1] (*n*=44)	
Pre Mean, SD	0.33, 0.55
Longitudinal Mean, SD	0.18, 0.65
Mean [of] Difference [SD]	-0.15
Effect Size (SE)	-0.25 (0.21)
Cohen's descriptor	Small (negative)
Grade 7/9[1] (*n*=68)	
Pre Mean, SD	0.26, 0.61
Longitudinal Mean, SD	0.66, 0.64
Mean [of] Difference [SD]	0.40
Effect Size (SE)	0.64 (0.18)
Cohen's descriptor	Medium
Grade 9/11[1] (*n*=31)	
Pre Mean, SD	0.62, 0.50
Longitudinal Mean, SD	0.87, 0.52
Mean [of] Difference [SD]	0.25
Effect Size (SE)	0.48 (0.26)
Cohen's descriptor	Medium

[1]Grade in Longitudinal follow-up

Part 3 – Comparison of Intervention and Non-Intervention groups

Table 8 shows the pre and longitudinal data for each grade in order to compare the degree of change for the two groups, Intervention and Non-Intervention. For the comparison of Grades 3 and 5, the Grade 3 groups were different from the outset in favour of the Intervention group but the difference decreased to a small extent by the time of the longitudinal survey. The effect size increased, however, because of the more consistent performances of students on the longitudinal survey (seen in the SD values). For the comparative change in Grades 5 and 7, the Grade 5s were not very different at the outset and the difference increased to a small extent in favour of the Intervention group over two years. For the change in Grades 7 and 9 over the two years, the Grade 7s were slightly different at the outset favouring the Non-Intervention group but the difference decreased to a small extent still in favour of the Non-Intervention group. The Grade 9s being compared were different from the outset in favour of the Non-Intervention group and the difference increased to a small extent in favour of that Non-Intervention group over the two years. This suggests that at mid to upper secondary level the Intervention did not result in a long-term effect for students staying in school that matched the Non-Intervention.

Table 8
Results for Comparisons Between the Intervention and Non-Intervention Groups

	Pre test	Longitudinal
Grade 3/5[1]		
Intervention Mean, SD (*n*=47)	-0.19, 0.72	0.50, 0.50
Non-Intervention Mean, SD (*n*=67)	-0.46, 0.83	0.28, 0.42
Mean Diff	0.27	0.22
Effect Size (SE)	0.34 (0.19)	0.48 (0.19)
Cohen's descriptor	Small	Medium
Grade 5/7[1]		
Intervention (*n*=58)	0.35, 0.44	0.28, 0.62
Non-Intervention (*n*=44)	0.33, 0.55	0.18, 0.65
Mean Diff	0.02	0.10
Effect Size (SE)	0.04 (0.20)	0.16 (0.20)
Cohen's descriptor	Very small	Small
Grade 7/9[1]		
Intervention (*n*=67)	0.00, 0.75	0.57, 0.83
Non-Intervention (*n*=68)	0.26, 0.61	0.66, 0.64
Mean Diff	-0.26	-0.09
Effect Size (SE)	-0.38 (0.17)	-0.12 (0.17)
Cohen's descriptor	Small	Small
Grade 9/11[1]		
Intervention (*n*=28)	0.38, 0.62	0.53, 0.79
Non-Intervention (*n*=31)	0.62, 0.50	0.87, 0.52
Mean Diff	-0.24	-0.34
Effect Size (SE)	-0.42 (0.26)	-0.51 (0.26)
Cohen's descriptor	Small	Medium

[1]Grade in Longitudinal follow-up

The change over two years for each of the four grade levels for each group, Intervention and Non-Intervention, is shown in Figure 2. Although the overall impression may be one of progression with grade, a comparison of the four Grade 5 means with the four Grade 7 means illustrates the lack of progress over these middle years in particular, even though the improvement after Grade 7 looks better than perhaps it should.

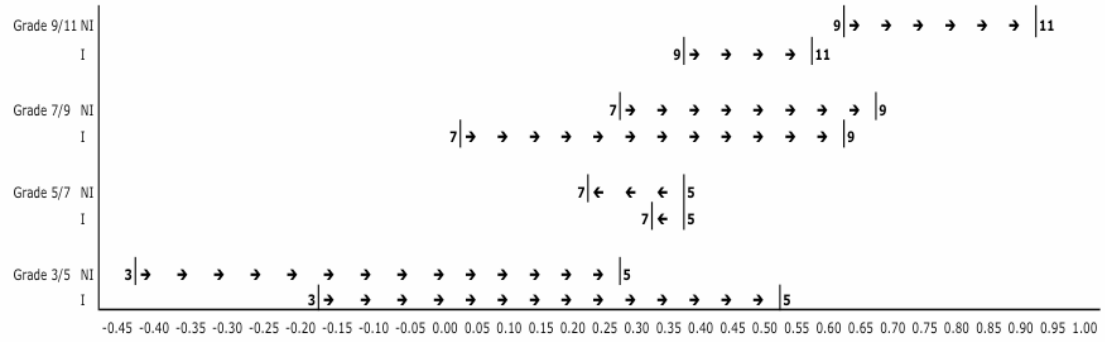


Figure 2. Difference in mean performance for Intervention (I) and Non-Intervention (NI) groups (arrows indicate the direction of change)

Discussion

Observations from the Rasch Analysis

The conclusions arising from the analysis presented in this paper point to some important lessons for studies that compare different cohorts and studies that undertake short-term interventions. The initial observation of a plateau in performance after Grade 5 for the common items (see Figure 1) was observed over a range of items within the survey. Watson and Kelly (2003a) found this levelling of performance on questions associated with the interpretation of a pictograph. In this case there may have been an issue of whether older students addressed the question as seriously as younger students did. Another item, dealing with the prediction of 60 outcomes from tossing a regular six-sided die (Watson & Kelly, 2003b), however, showed similar results, a plateau from Grade 5. Students' ability to suggest appropriate variation in outcomes did not improve after this point. In providing definitions for statistical terms, Watson and Kelly (2003c) found that for the word "sample" there was improvement from Grade 3 to 5 but little after that. For "random" and "variation," asked only to Grades 7 and 9, there was little difference in performance over the two years.

The greatest short-term improvement in performance for the Intervention group in this study occurred for Grade 7 students, judged by Cohen's criteria to be a medium effect. For Grades 3 and 5 the improvement was judged as small. For some of the specific tasks in the survey, performance by students reflected this overall. For example, for the question on predicting dice outcomes, Grade 3 and 9 improved marginally, Grade 7 improved the most, and Grade 5 students not at all. The issue of Grade 7 improvement, as noted earlier, is tempered by the lower starting mean for this group. The Grade 9 drop on the post-test in the Intervention group for the students still in the study after two years (n=28) may be related to the small sample size, even though these students proceeded to Grade 11. For the larger subgroup of students who completed the post test (n=90) the result was somewhat better, reflecting what was reported in Watson & Kelly (2002c).

That the Non-Intervention group showed similar levels of change over two years is interesting in terms of the larger educational milieu. Differences were either small, very small, or medium but in opposite directions for two different grades. Of greatest concern to educators is the slightly diminished performance from Grade 5 to Grade 7 in both Intervention and Non-Intervention groups. This middle school slump, also noted in Figure 1, has been reported elsewhere (e.g., Hill, Rowe, Holmes-Smith, & Russell, 1996; Callingham & McIntosh, 2002). As noted earlier all students changed schools in this time, albeit in the same local neighbourhood. It is unknown if the drop of about 1/3 in the sample size over two years might include the movement of some more of the able students to the private school sector. As evidence accumulates from various studies, more attention needs to be paid to the issue across these years.

The similar long term performance of the Intervention and Non-Intervention high schools may be explained to some extent by the attendance of several Non-Intervention teachers in a Quality Teacher Program, including sessions on chance and data led by the first author. Although not expected at the beginning of the project, or ideal for the project results, it was not ethically possible to exclude the teachers from the program. As well there may be concerns that the Intervention schools, having had a program specifically aimed at chance and data in the first year of the project may

have neglected these topics in the subsequent years. It may also be that the strength of the intervention either as implemented by teachers, or as suggested by the research team, was not sufficient to sustain the improvement found in the post test. Overall it must be acknowledged that more must be done to achieve the desired improvements from the standpoint of the chance and data curriculum.

Comparison of Results with Conventional Analysis

Watson and Kelly (in press) considered the two-year longitudinal data from this study in an alternative fashion using conventional analysis (sum of raw scores on items with *t*-tests). For the total scores on the surveys the longitudinal change was significant with $p<0.0001$ for Grade 3 and Grade 7 in both Intervention and Non-Intervention groups and with $p<0.005$ for Grade 9 in the Non-Intervention group. For Grades 5 and 9 in the Intervention group the change was positive but non-significant, whereas for Grade 5 in the Non-Intervention group the change was negative and non-significant. In the light of the debate about the use of significance tests versus the consideration of effect size (Cohen, 1969) as used earlier in this paper with the results of the Rasch analysis, it is of interest to view the effect sizes for the outcomes from the conventionally scored surveys. These are given for each grade in Table 9 in the same format as earlier in Table 6 and 7.

Table 9
Results for Comparisons of Differences over Two Years within Intervention and Non-Intervention Groups using Conventional Scoring

	Intervention[1]	Non-Intervention[1]
Grade 3/5[2]	(*n*=56)	(*n*=67)
Pre-test Mean, SD	24.09, 9.32	21.64, 8.78
Post or Long Mean, SD	35.14, 7.98	32.01, 7.17
Mean Diff	11.05	10.37
Effect Size (SE)	1.26 (0.21)	1.29 (0.19)
Cohen's descriptor	Large	Large
Grade 5/7[2]	(*n*=61)	(*n*=44)
Pre-test Mean, SD	37.11, 9.29	36.50, 10.2
Post or Long Mean, SD	38.11, 12.65	35.48, 11.6
Mean Diff	1.00	-1.02
Effect Size (SE)	0.09 (0.18)	-0.09 (0.21)
Cohen's descriptor	Very small	Very small
Grade 7/9[2]	(*n*=72)	(*n*=68)
Pre-test Mean, SD	42.31, 17.8	49.94, 16.8
Post or Long Mean, SD	58.63, 22.7	62.68, 18.4
Mean Diff	16.32	12.74
Effect Size (SE)	0.80 (0.17)	0.72 (0.18)
Cohen's descriptor	Large	Medium
Grade 9/11[2]	(*n*=30)	(*n*=31)
Pre-test Mean, SD	54.40, 17.5	62.16, 22.8
Post or Long Mean, SD	59.03, 22.5	70.06, 16.4
Mean Diff	4.63	7.90
Effect Size (SE)	0.23 (0.26)	0.40 (0.26)
Cohen's descriptor	Small	Small

[1]Different sample sizes are because complete post data were not required in the Intervention analysis (data from Watson & Kelly, in press)
[2]Grade in Longitudinal follow-up

In considering the difference between the two analyses, it should be noted that the data used for the calculations in Table 9 reflect only the results for the subset of common items completed by students in the initial grade. Grade 5 students, for example, when being compared with their performances as Grade 3 students two years earlier, were only scored on the items completed in Grade 3, not the extra items completed in Grade 5. In the Rasch analysis reported earlier in this paper the students in later grades had scaled scores that reflected all of the items completed in that grade, providing more comprehensive information on the students as they matured and were able to attempt more demanding questions.

To clarify the comparison of the effect sizes on the left of Table 9 with those on the right of Table 6 and on the right of Table 9 with those in Table 7, a summary is presented in Table 10. For Grade 7/9 in the Intervention group where the Rasch effect is medium and the conventional effect is large, in fact the two effect values (0.72 and 0.80) are both close to the boundary described by Cohen (1969) (0.75). Similarly for Grade 9/11 in the Non-Intervention group, the Rasch effect is medium (0.48) whereas the conventional effect is small (0.40) and the boundary is 0.44. Otherwise the effect sizes are nearly the same using the two different approaches.

Table 10
Comparison of Effect Sizes for Different Methods of Analysis

	Intervention		Non-Intervention	
	Rasch	Conventional	Rasch	Conventional
Grade 3/5[1]	1.10 Large	1.26 Large	1.12 Large	1.29 Large
Grade 5/7[1]	-0.13 Very small (negative)	0.09 Very small (positive)	-0.25 Small (negative)	-0.09 Very Small (negative)
Grade 7/9[1]	0.72 Medium	0.80 Large	0.64 Medium	0.72 Medium
Grade 9/11[1]	0.21 Small	0.23 Small	0.48 Medium	0.40 Small

[1]Grade in Longitudinal follow-up

Conclusion

Two types of issues arise in this study. One is related to the method of measurement used to assess student understanding in the context of the chance and data curriculum. Using Partial Credit Rasch analysis has allowed information from all questions answered by students to be used making comparisons across groups and time. The large set of common items (n=21) answered by all students provides a sound foundation for the linking with other items and therefore a more comprehensive analysis. In the future this analysis will be the basis, using anchored items, for the comparison of student performance with respect to curriculum intentions across the years 1993 to 2003. This and further descriptive analysis of individual items (e.g., Watson & Kelly, 2003a) will assist in future planning for curriculum and teaching.

Perhaps the most significant educational finding from the analysis is the levelling of performance over the middle years of schooling. That this occurred in the initial data set as well as in the two-year data for Grades 5/7, both for the Intervention and Non-Intervention groups, gives evidence that it is a real effect not associated with

the cohorts involved or the intervention used in this research. Although there might be some mitigating circumstances (like transfers to other educational systems), it would appear that education systems need to take heed of a potentially serious problem.

Acknowledgements

This research was funded by Australian Research Council grants (No. A00000716 and No. DP0208607).

References

Adams, R. J., & Khoo, S. T. (1996). *Quest: Interactive item analysis system. Version 2.1* [Computer software]. Melbourne: Australian Council for Educational Research.

Australian Education Council. (1991). *A national statement on mathematics for Australian schools.* Carlton, VIC: Author.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, N J: Lawrence Erlbaum.

Callingham, R., & McIntosh, A. (2002). Mental computation competence across years 3 to 10. In B. Barton, K. C. Irwin, M. Pfannkuch & M. O. J. Thomas, (Eds.), *Mathematics education in the South Pacific: Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland* (pp. 155-163). Sydney: MERGA.

Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.

Cobb, P. (1999). Individual and collective mathematical development: The case for statistical data analysis. *Mathematical Thinking and Learning, 1*, 5-43.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382-393.

Department of Education and the Arts Tasmania (1992). *Mathematics guidelines K-8*. Hobart: Curriculum Services Branch.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? An exploratory research study. *Educational Studies in Mathematics, 15*, 1-24.

Green, D. (1991). A longitudinal study of pupils' probability concepts. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics: School and general issues* (Vol. 1, pp. 320-328). Voorburg, The Netherlands: International Statistical Institute.

Green, D. (1993). Data analysis: What research do we need? In L. Pereira-Mendoza (Ed.), *Introducing data analysis in the schools: Who should teach it?* (pp. 219-239). Voorburg, The Netherlands: International Statistical Institute.

Green, D. R. (1986). Children's understanding of randomness: Report of a survey of 1600 children aged 7-11 years. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 287-291). Victoria, BC: The Organizing Committee, ICOTS2.

Hill, P. W., Rowe, K. J., Holmes-Smith, P., & Russell, V. J. (1996). *The Victorian Quality Schools Project: A study of school and teacher effectiveness. Report (Volume 1).* Melbourne: Centre for Applied Educational Research, University of Melbourne.

Izard, J. F. (2004, March). *Best practice in assessment for learning*. Paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies on *Redefining the Roles of Educational Assessment*, South Pacific Board for Educational Assessment, Nadi, Fiji.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, *26*(1), 20-39.

Moore, D. S. (1990). Uncertainty. In L.A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published 1960, Copenhagen: Danish Institute for Educational Research).

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph and K. Carr (Eds.), *People in mathematics education*: *Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia*, MERGA, Rotorua, NZ, pp. 6-22.

Shaughnessy, J. M., Canada, D., & Ciancetta, M. (2003). Middle school students' thinking about variability in repeated trials: A cross-task comparison. In N.A. Pateman, B.J. Dougherty, & J.T. Zilliox (Eds.), *Proceedings of the 27th conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th conference of PME-NA* (Vol. 4, pp. 159-165). Honolulu, HI: Center for Research and Development Group, University of Hawaii.

Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*, 247-275.

Watson, J. M., & Kelly, B. A. (2002a). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching*

Statistics: Developing a statistically literate society, Cape Town, South Africa. Voorburg, The Netherlands: International Statistical Institute.

Watson, J. M., & Kelly, B. A. (2002b). Grade 5 students' appreciation of variation. In A. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 385-392). Norwich, UK: University of East Anglia.

Watson, J. M., & Kelly, B. A. (2002c). Variation as part of chance and data in grades 7 and 9. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific: Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia, Auckland*. (Vol. 2, pp. 682-689). Sydney, NSW: MERGA.

Watson, J. M., & Kelly, B. A. (2003a). Inference from a pictograph: Statistical literacy in action. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity: Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia Conference, Geelong* (pp. 720-727). Sydney, NSW: MERGA.

Watson, J. M., & Kelly, B. A. (2003b). Predicting dice outcomes: The dilemma of expectation versus variation. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity: Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia Conference, Geelong* (pp. 728-735). Sydney, NSW: MERGA.

Watson, J. M., & Kelly, B. A. (2003c). The vocabulary of statistical literacy. In *Educational research, risks, & dilemmas: Proceedings of the joint conferences of the New Zealand Association for Research in Education and the Australian Association for Research in Education* [CD-ROM]. Auckland, New Zealand.

Watson, J. M., & Kelly, B. A. (2004). A two-year study of students' appreciation of variation in the chance and data curriculum. In I. Putt, R. Faragher, & M. McLean (Eds.), *Mathematics education for the third millennium: Towards 2010*: *Proceedings of the 27th Annual Conference of the Mathematics Education Research Group of Australasia, Townsville* (Vol. 2, pp. 573-580). Sydney, NSW: MERGA.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*, 1-29.

Appendix A

Item Fit

Items 1 to 21 2000 data (Run No 1) [Items in common]

Item Estimates (Thresholds) In input Order all on all (N = 738 L = 21 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	DIE7	1455	2952	-1.72	-0.01	0.37	1.89		1.02	1.03	0.4	0.5
				.19	.14	.13	.16					
2	**DIE2**	**984**	**2952**	**-0.47**	**0.39**	**1.10**	**2.16**		**1.22**	**1.43**	**4.4**	**5.7**
				.13	**.13**	**.15**	**.20**					
3	SP1	1138	1476	-1.94	-0.35				0.96	0.90	-0.7	-1.3
				.22	.16							
4	TRV1	698	738	-2.91					0.94	1.02	-0.4	0.2
				.17								
5	TRV2	579	738	-1.26					0.94	0.93	-1.1	-0.9
				.10								
6	TRV3	1137	1476	-1.13	-0.54				0.99	0.98	-0.1	-0.2
				.17	.16							
7	**TRV4**	**1418**	**3690**	**-1.69**	**-0.15**	**0.78**	**1.38**	**2.28**	**1.30**	**1.32**	**5.3**	**4.6**
				.16	**.13**	**.15**	**.19**	**.24**				
8	TRV5	665	1476	-1.84	3.39				0.95	0.89	-0.6	-1.2
				.22	.29							
9	**TRV6**	**1052**	**3690**	**-1.91**	**0.84**	**1.01**	**1.23**	**3.34**	**1.25**	**1.56**	**3.8**	**5.4**
				.16	**.15**	**.16**	**.17**	**.40**				
10	MVE1	695	2952	-0.50	1.14	1.73	2.44		1.00	1.06	0.1	0.8
				.13	.18	.22	.31					
11	MVE2	503	2214	0.41	1.00	2.15			0.92	0.91	-1.7	-0.9
				.16	.14	.20						
12	MVE3	628	2214	0.19	0.55	3.99			0.84	0.77	-4.2	-2.9
				.13	.15	.42						
13	MVE4	784	2214	0.00	0.53	1.35			0.93	0.88	-1.7	-1.5
				.13	.12	.13						
14	MVE7	774	2214	-0.25	0.78	1.22			1.07	1.18	1.4	2.4
				.16	.14	.14						
15	SMP4	664	2214	-0.19	0.96	2.01			0.87	0.83	-2.8	-2.6
				.13	.17	.17						
16	BOX9	1060	2214	-0.31	-0.12	1.44			0.94	0.96	-1.4	-0.4
				.13	.12	.13						
17	TBL1	589	738	-1.35					0.94	0.83	-1.1	-2.1
				.10								
18	TBL2	597	738	-1.42					0.89	0.77	-2.0	-2.8
				.10								
19	TBL3	1106	1476	-0.94	-0.47				0.95	0.93	-0.9	-0.7
				.16	.15							
20	TBL4	528	738	-0.87					0.93	0.90	-1.6	-1.6
				.09								

*****Output Continues****

```
Items 1 to 21 2000 data (Run No 1) [Items in common]
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order  all on all (N = 738 L = 21 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------------
21  TBL5           |  1017 2952 |  -0.44    0.52    0.62    2.96        |   1.05  1.11   1.1   1.5
                   |            |     .13     .13     .14     .27        |
                   |            |                                       |
-------------------------------------------------------------------------------------------------------
Mean               |            |   0.00                                |   1.00  1.01  -0.2   0.1
SD                 |            |   1.17                                |   0.12  0.21   2.3   2.5
=======================================================================================================
```

Comment: Three items (shown **bold**) out of 21 have possible misfit.

```
Items 1 to 21 2000 data (Run No 1) [Items in common]
-------------------------------------------------------------------------------------------------------
Item Fit all on all (N = 738 L = 21 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
INFIT
 MNSQ  0.50      0.56      0.63      0.71      0.83      1.00      1.20      1.40      1.60
---------+---------+---------+---------+---------+---------+---------+---------+---------+-----
  1 DIE7                                  .               |*              .
  2 DIE2                                  .               |           *   .
  3 SP1                                   .             * |               .
  4 TRV1                                  .            *  |               .
  5 TRV2                                  .            *  |               .
  6 TRV3                                  .               *              .
  7 TRV4                                  .               |              *
  8 TRV5                                  .             * |               .
  9 TRV6                                  .               |            * .
 10 MVE1                                  .               *              .
 11 MVE2                                  .          *    |               .
 12 MVE3                                  .     *         |               .
 13 MVE4                                  .           *   |               .
 14 MVE7                                  .               |  *            .
 15 SMP4                                  .      *        |               .
 16 BOX9                                  .            *  |               .
 17 TBL1                                  .            *  |               .
 18 TBL2                                  .         *     |               .
 19 TBL3                                  .            *  |               .
 20 TBL4                                  .           *   |               .
 21 TBL5                                  .               |  *            .
=======================================================================================================
```

Comment: All items have acceptable item fit. *Decision*: All items have acceptable fit.

```
Items 1 to 25 2000 data (Run No B1) [Grades 3 and 5]
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order all on all (N = 359 L = 25 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------------
1   DIE7           |   561 1436 |  -1.72   -0.01    0.37    1.89        |   0.87  0.87  -2.1  -1.5
                   |            |      **      **      **      **        |
                   |            |                                       |
2   DIE2           |   397 1436 |  -0.47    0.39    1.10    2.16        |   1.09  1.05   1.3   0.5
                   |            |      **      **      **      **        |
                   |            |                                       |
3   SP1            |   457  718 |  -1.94   -0.35                        |   1.03  0.98   0.4  -0.2
                   |            |      **      **                       |
                   |            |                                       |
4   TRV1           |   334  359 |  -2.91                                |   1.02  1.27   0.2   1.1
                   |            |      **                               |
                   |            |                                       |
=======================================================================================================
                                                *****Output Continues****
```

Items 1 to 25 2000 data (Run No B1) [Grades 3 and 5]

Item Estimates (Thresholds) In input Order all on all (N = 359 L = 25 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
5	TRV2	271	359	-1.26 **					0.88	0.85	-1.8	-1.5
6	TRV3	518	718	-1.13 **	-0.54 **				0.95	0.89	-0.6	-0.9
7	TRV4	676	1795	-1.69 **	-0.15 **	0.78 **	1.38 **	2.28 **	1.03	1.03	0.4	0.4
8	TRV5	307	718	-1.84 **	3.39 **				1.04	0.94	0.4	-0.5
9	TRV6	478	1436	-1.91 **	0.85 **	1.05 **	1.29 **		1.12	1.30	1.4	2.2
10	MVE1	306	1436	-0.50 **	1.14 **	1.73 **	2.44 **		0.92	0.99	-0.9	-0.1
11	MVE2	222	1077	0.41 **	1.00 **	2.15 **			0.90	1.00	-1.3	0.1
12	MVE3	249	718	0.19 **	0.58 **				0.80	0.73	-4.1	-2.6
13	MVE4	295	1077	0.00 **	0.53 **	1.35 **			0.81	0.76	-3.2	-2.3
14	**MVE7**	**442**	**1077**	**-0.25 **	**0.78 **	**1.22 **			**1.43**	**1.64**	**5.5**	**5.4**
15	SMP4	249	1077	-0.19 **	0.96 **	2.01 **			0.69	0.65	-4.8	-4.0
16	BOX9	549	1077	-0.31 **	-0.12 **	1.44 **			0.87	0.99	-2.2	0.0
17	TBL1	325	359	-1.35 **					0.59	0.54	-6.5	-5.0
18	TBL2	333	359	-1.42 **					0.45	0.37	-9.0	-7.4
19	TBL3	606	718	-0.94 **	-0.47 **				0.51	0.49	-8.3	-5.2
20	TBL4	279	359	-0.87 **					0.84	0.84	-3.2	-2.0
21	TBL5	566	1436	-0.44 **	0.52 **	0.62 **	2.96 **		0.93	1.09	-1.1	0.9
22	SP2A	468	1077	-1.56 .25	0.42 .22	1.84 .29			0.76	0.76	-3.5	-2.8
23	SP3A	396	1077	-0.36 .17	0.39 .19	1.10 .20			0.82	0.76	-3.0	-2.6

*****Output Continues****

```
Items 1 to 25 2000 data (Run No B1) [Grades 3 and 5]
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order all on all (N = 359 L = 25 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------------
24  SP4A           |   172  359 |   0.10                                |   0.90  0.88  -3.3  -1.7
                   |            |     .11                               |
                   |            |                                       |
25  SP5A           |   538 1074 |  -0.30   -0.10    0.56                |   1.18  1.24   2.8   1.8
                   |            |     .19     .18     .17               |
                   |            |                                       |
-------------------------------------------------------------------------------------------------------
Mean               |            |  -0.04                                |   0.90  0.92  -1.9  -1.1
SD                 |            |   1.01                                |   0.21  0.27   3.3   2.7
=======================================================================================================
```

Comment: One item out of 25 (shown **bold**) has possible misfit.

```
Items 1 to 25 2000 data (Run No B1) [Grades 3 and 5]
-----------------------------------------------------------------------------
Item Fit all on all (N = 359 L = 25 Probability Level=0.50)
-----------------------------------------------------------------------------------------
INFIT
 MNSQ  0.40      0.45      0.53      0.63      0.77      1.00      1.30      1.60      1.90
---------+---------+---------+---------+---------+---------+---------+---------+---------+------
  1 DIE7                                    .    *    |         .
  2 DIE2                                    .         |  *      .
  3 SP1                                     .         |*        .
  4 TRV1                                    .         |*        .
  5 TRV2                                    .    *    |         .
  6 TRV3                                    .        * |        .
  7 TRV4                                    .         |*        .
  8 TRV5                                    .         |*        .
  9 TRV6                                    .         |   *     .
 10 MVE1                                    .       * |         .
 11 MVE2                                    .      *  |         .
 12 MVE3                                    .*        |         .
 13 MVE4                                    . *       |         .
 14 MVE7                                    .         |         .   *
 15 SMP4                                 *  .         |         .
 16 BOX9                                    .    *    |         .
 17 TBL1                              *     .         |         .
 18 TBL2          *                         .         |         .
 19 TBL3                    *               .         |         .
 20 TBL4                                    .   *     |         .
 21 TBL5                                    .      *  |         .
 22 SP2A                                   *.         |         .
 23 SP3A                                    .  *      |         .
 24 SP4A                                    .     *   |         .
 25 SP5A                                    .         |     *   .
=========================================================================================
```

Comment: Pattern of item fit subject to constraints due to smaller sample and restricted range of achievement. *Decision*: New items (shown **bold**) have acceptable fit.

```
Items 1 to 29 2000 data (Run No B2) [Grade 5]
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order all on all (N = 183 L = 29 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------------
1   DIE7           |   356  732 |  -1.72   -0.01    0.37    1.89         |   0.84  0.82  -1.9  -1.6
                   |            |     **      **      **      **         |
                   |            |                                       |
2   DIE2           |   261  732 |  -0.47    0.39    1.10    2.16         |   1.10  1.08   1.1   0.7
                   |            |     **      **      **      **         |
                   |            |                                       |
=======================================================================================================
                                                *****Output Continues****
```

Items 1 to 29 2000 data (Run No B2) [Grade 5]

Item Estimates (Thresholds) In input Order all on all (N = 183 L = 29 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
3	SP1	290	366	-1.94 **	-0.35 **				0.82	0.77	-1.5	-1.6
4	TRV1	177	183	-2.91 **					0.88	1.57	-0.3	1.3
5	TRV2	157	183	-1.26 **					0.80	0.78	-1.7	-1.4
6	TRV3	298	366	-1.13 **	-0.54 **				1.02	0.99	0.2	0.0
7	TRV4	414	915	-1.69 **	-0.15 **	0.78 **	1.38 **	2.28 **	1.14	1.19	1.4	1.5
8	TRV5	172	366	-1.84 **	3.39 **				0.95	0.85	-0.2	-0.8
9	TRV6	283	732	-1.91 **	0.85 **	1.05 **	1.29 **		1.06	1.25	0.6	1.5
10	MVE1	184	732	-0.50 **	1.14 **	1.73 **	2.44 **		0.89	0.93	-0.9	-0.4
11	MVE2	149	549	0.41 **	1.00 **	2.15 **			0.80	0.85	-2.4	-1.1
12	MVE3	166	366	0.19 **	0.58 **				0.86	0.84	-2.3	-1.5
13	MVE4	200	549	0.00 **	0.53 **	1.35 **			0.90	0.90	-1.3	-0.9
14	MVE7	267	549	-0.25 **	0.78 **	1.22 **			1.23	1.25	2.6	2.0
15	SMP4	173	549	-0.19 **	0.96 **	2.01 **			0.68	0.67	-4.0	-3.1
16	BOX9	319	549	-0.31 **	-0.12 **	1.44 **			0.72	0.72	-3.5	-2.4
17	TBL1	174	183	-1.35 **					0.45	0.38	-5.2	-4.6
18	TBL2	179	183	-1.42 **					0.29	0.24	-7.2	-6.1
19	TBL3	344	366	-0.94 **	-0.47 **				0.43	0.35	-5.7	-4.6
20	TBL4	158	183	-0.87 **					0.71	0.74	-3.4	-2.1
21	TBL5	367	732	-0.44 **	0.52 **	0.62 **	2.96 **		0.84	0.84	-2.1	-1.3
22	SP2A	289	549	-2.63 .63	0.24 .30	2.30 .40			0.82	0.82	-1.9	-1.5
23	SP3A	255	549	-0.50 .28	0.46 .25	1.43 .30			0.83	0.82	-2.1	-1.7

*****Output Continues****

Items 1 to 29 2000 data (Run No B2) [Grade 5]

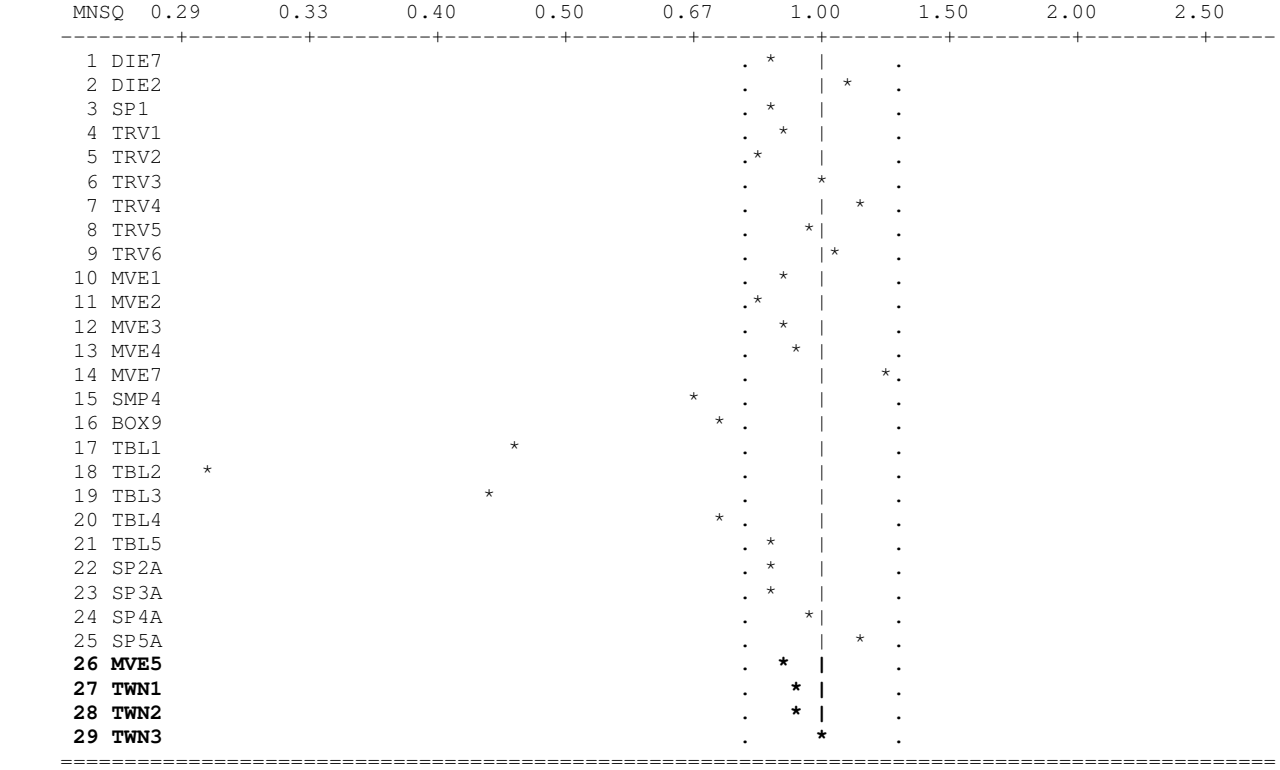
Item Estimates (Thresholds) In input Order all on all (N = 183 L = 29 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
24	SP4A	108	183	0.01					0.94	0.93	-1.3	-0.7
				.16								
25	SP5A	304	549	-0.09	0.05	0.86			1.14	1.22	1.7	1.7
				.23	.25	.22						
26	MVE5	183	549	0.03	0.68	3.14			0.88	0.86	-1.7	-1.3
				.27	.27	.57						
27	TWN1	215	549	-0.19	0.68	1.53			0.92	0.91	-0.9	-0.7
				.25	.25	.31						
28	TWN2	180	549	0.17	0.77	1.62			0.90	0.85	-1.1	-1.2
				.25	.26	.31						
29	TWN3	192	549	-0.09	0.81	1.87			1.00	1.00	0.1	0.0
				.25	.27	.34						
Mean				0.09					0.86	0.88	-1.6	-1.1
SD				1.00					0.21	0.27	2.2	1.9

Comment: All items have acceptable fit.

Items 1 to 29 2000 data (Run No B2) [Grade 5]

Item Fit all on all (N = 183 L = 29 Probability Level=0.50)



Comment: Pattern of item fit subject to more constraints due to even smaller sample and restricted range of achievement. *Decision*: New items (shown **bold**) have acceptable fit.

Items 1 to 21 and 26 to 49 2000 data (Run No C1) [Grades 7 and 9]

Item Estimates (Thresholds) In input Order all on all (N = 379 L = 45 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	DIE7	897	1516	-1.72 **	-0.01 **	0.37 **	1.89 **		1.11	1.15	1.7	1.7
2	**DIE2**	**584**	**1516**	**-0.47 ****	**0.39 ****	**1.10 ****	**2.16 ****		**1.24**	**1.64**	**3.5**	**5.8**
3	SP1	682	758	-1.94 **	-0.35 **				0.69	0.62	-4.5	-4.3
4	TRV1	364	379	-2.91 **					0.64	0.55	-2.0	-1.9
5	TRV2	307	379	-1.26 **					0.95	0.93	-0.7	-0.6
6	TRV3	620	758	-1.13 **	-0.54 **				0.96	0.98	-0.4	-0.1
7	**TRV4**	**742**	**1895**	**-1.69 ****	**-0.15 ****	**0.78 ****	**1.38 ****	**2.28 ****	**1.48**	**1.50**	**5.9**	**4.9**
8	TRV5	357	758	-1.84 **	3.39 **				0.74	0.72	-2.7	-2.4
9	TRV6	574	1895	-1.91 **	0.64 **	0.77 **	0.87 **	1.16 **	1.10	1.50	1.3	3.6
10	MVE1	391	1516	-0.50 **	1.14 **	1.73 **	2.44 **		1.04	1.12	0.5	1.2
11	MVE2	281	1137	0.41 **	1.00 **	2.15 **			0.96	0.89	-0.6	-0.9
12	MVE3	378	1137	-0.02 **	0.22 **	0.87 **			0.66	0.63	-6.8	-3.5
13	MVE4	489	1137	0.00 **	0.53 **	1.35 **			0.97	0.90	-0.6	-0.9
14	MVE7	331	1137	-0.25 **	0.78 **	1.22 **			0.76	0.80	-4.1	-2.2
15	SMP4	415	1137	-0.19 **	0.96 **	2.01 **			1.04	1.08	0.6	0.8
16	BOX9	511	1137	-0.31 **	-0.12 **	1.44 **			1.16	1.12	2.5	1.0
17	TBL1	265	379	-1.35 **					1.36	1.27	4.0	2.1
18	TBL2	263	379	-1.42 **					1.43	1.34	4.5	2.6
19	**TBL3**	**499**	**758**	**-0.94 ****	**-0.47 ****				**1.64**	**1.66**	**6.9**	**4.3**
20	TBL4	248	379	-0.87 **					1.13	1.09	2.1	1.0
21	TBL5	452	1516	-0.44 **	0.52 **	0.62 **	2.96 **		1.28	1.27	4.3	2.6

*****Output Continues****

Items 1 to 21 and 26 to 49 2000 data (Run No C1) [Grades 7 and 9]

Item Estimates (Thresholds) In input Order all on all (N = 379 L = 45 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
26	MVE5	385	1137	0.03	0.68	3.14			1.15	1.09	2.7	0.9
				**	**	**						
27	TWN1	528	1137	-0.19	0.68	1.53			1.30	1.29	4.5	2.9
				**	**	**						
28	TWN2	443	1137	0.17	0.77	1.62			1.29	1.33	4.3	2.7
				**	**	**						
29	TWN3	492	1137	-0.09	0.81	1.87			1.48	1.53	6.6	4.8
				**	**	**						
30	SP6	278	379	-0.87					1.04	1.02	0.6	0.2
				.12								
31	SP7	128	379	0.90					0.98	0.96	-0.5	-0.3
				.11								
32	SP8	112	379	1.10					0.92	0.87	-1.7	-1.3
				.12								
33	SP9	35	379	2.52					0.93	0.73	-0.5	-1.3
				.18								
34	SP10	326	1137	0.34	0.85	1.24			1.19	1.29	2.6	2.1
				.19	.18	.19						
35	SP11A	782	1137	-1.41	-0.45	0.41			1.18	1.27	2.5	2.6
				.23	.19	.21						
36	SP11B	554	1137	-0.63	0.47	0.71			1.13	1.12	2.2	1.3
				.19	.18	.17						
37	SP11C	683	1137	-1.16	-0.19	0.87			0.98	0.97	-0.3	-0.3
				.22	.20	.17						
38	MVE6	252	1137	0.19	1.52	2.00			0.85	0.81	-1.8	-1.7
				.19	.26	.31						
39	BT1A	305	1137	0.44	0.56	3.35			0.88	0.77	-2.3	-1.2
				.19	.18	.43						
40	BT1B	169	1137	0.72	1.84	2.08			0.92	0.82	-0.7	-1.4
				.22	.34	.35						
41	RAN3	304	1137	0.22	0.83	3.85			0.89	0.83	-2.1	-1.7
				.19	.20	.55						
42	VAR	290	1137	0.34	0.85	2.42			0.88	0.83	-1.9	-1.5
				.19	.20	.32						
43	AVG1	244	1137	0.56	0.97	2.45			0.93	0.89	-1.0	-0.7
				.19	.21	.32						
44	DRG1	58	379	1.95					0.89	0.75	-1.1	-1.6
				.15								
45	M4DR	122	758	1.13	1.69				0.87	0.79	-1.4	-1.2
				.25	.26							
46	SP2B	717	1137	-3.13	-0.82	1.73			1.00	0.99	0.0	-0.1
				.47	.25	.24						

*****Output Continues****

Items 1 to 21 and 26 to 49 2000 data (Run No C1) [Grades 7 and 9]

Item Estimates (Thresholds) In input Order all on all (N = 379 L = 45 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
47	SP3B	610	1137	-0.84	0.00	1.16			1.12	1.18	1.9	2.0
				.19	.19	.19						
48	SP4B	278	379	-0.87					1.04	1.06	0.8	0.6
				.12								
49	SP5B	603	1137	-0.38	-0.24	1.37			1.22	1.26	3.2	1.9
				.19	.20	.19						
Mean				0.36					1.05	1.05	0.7	0.5
SD				1.08					0.22	0.28	3.0	2.3

Comment: Three items (shown **bold**) out of 45 have possible misfit.

```
Items 1 to 21 and 26 to 49 2000 data (Run No C1) [Grades 7 and 9]
-------------------------------------------------------------------------------------------
Item Fit all on all (N = 379 L = 45 Probability Level=0.50)
-------------------------------------------------------------------------------------------
INFIT
 MNSQ  0.50      0.56      0.63      0.71      0.83      1.00      1.20      1.40      1.60
---------+---------+---------+---------+---------+---------+---------+---------+---------+---
  1 DIE7                                    .              |    *         .
  2 DIE2                                    .              |           *  .
  3 SP1                               *     .              |              .
  4 TRV1                         *          .              |              .
  5 TRV2                                    .           *  |              .
  6 TRV3                                    .            * |              .
  7 TRV4                                    .              |              .        *
  8 TRV5                                *   .              |              .
  9 TRV6                                    .              |    *         .
 10 MVE1                                    .              | *            .
 11 MVE2                                    .            * |              .
 12 MVE3                          *         .              |              .
 13 MVE4                                    .            * |              .
 14 MVE7                                 *  .              |              .
 15 SMP4                                    .              | *            .
 16 BOX9                                    .              |       *      .
 17 TBL1                                    .              |              .  *
 18 TBL2                                    .              |              .      *
 19 TBL3                                    .              |              .                *
 20 TBL4                                    .              |     *        .
 21 TBL5                                    .              |          *.
 26 MVE5                                    .              |       *      .
 27 TWN1                                    .              |              *
 28 TWN2                                    .              |              *
 29 TWN3                                    .              |              .        *
 30 SP6                                     .              | *            .
 31 SP7                                     .            *|              .
 32 SP8                                     .          *  |              .
 33 SP9                                     .           *  |              .
 34 SP10                                    .              |         *    .
 35 SP11A                                   .              |         *    .
 36 SP11B                                   .              |       *      .
 37 SP11C                                   .            *|              .
 38 MVE6                                    .     *        |              .
 39 BT1A                                    .       *      |              .
 40 BT1B                                    .          *   |              .
 41 RAN3                                    .        *     |              .
 42 VAR                                     .       *      |              .
 43 AVG1                                    .          *   |              .
 44 DRG1                                    .        *     |              .
 45 M4DR                                    .       *      |              .
 46 SP2B                                    .              *              .
 47 SP3B                                    .              |     *        .
 48 SP4B                                    .              | *            .
 49 SP5B                                    .              |          *   .
===========================================================================================
```

Comment: Pattern of item fit subject to constraints due to smaller sample and restricted range of achievement. *Decision*: New items (shown **bold**) have acceptable fit.

```
Items 1 to 21 and 26 to 50 2000 data (Run No C2) [Grade 9]
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order all on all (N = 193 L = 46 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------------
1   DIE7           |   487  772 |  -1.72   -0.01    0.37    1.89         |   1.29  1.29   3.2   2.3
                   |            |      **      **      **      **        |
                   |            |                                       |
2   DIE2           |   305  772 |  -0.47    0.39    1.10    2.16         |   1.09  1.75   1.0   4.8
                   |            |      **      **      **      **        |
                   |            |                                       |
=======================================================================================================
                                                *****Output Continues****
```

Items 1 to 21 and 26 to 50 2000 data (Run No C2) [Grade 9]

Item Estimates (Thresholds) In input Order all on all (N = 193 L = 46 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
3	SP1	354	386	-1.94 **	-0.35 **				0.66	0.59	-3.5	-3.3
4	TRV1	186	193	-2.91 **					0.70	0.63	-1.0	-1.0
5	TRV2	163	193	-1.26 **					0.87	0.83	-1.2	-1.0
6	TRV3	332	386	-1.13 **	-0.54 **				0.94	0.93	-0.4	-0.3
7	**TRV4**	**350**	**965**	**-1.69 ****	**-0.15 ****	**0.78 ****	**1.38 ****	**2.28 ****	**1.64**	**1.60**	**5.4**	**4.1**
8	TRV5	183	386	-1.84 **	3.39 **				0.84	0.84	-1.1	-0.9
9	TRV6	305	965	-1.91 **	0.64 **	0.77 **	0.87 **	1.16 **	0.95	1.16	-0.4	1.0
10	MVE1	232	772	-0.50 **	1.14 **	1.73 **	2.44 **		1.12	1.23	1.1	1.5
11	MVE2	172	579	0.41 **	1.00 **	2.15 **			1.06	1.01	0.7	0.1
12	MVE3	221	579	-0.02 **	0.22 **	0.87 **			0.60	0.60	-6.0	-2.7
13	MVE4	288	579	0.00 **	0.53 **	1.35 **			0.99	0.95	0.0	-0.3
14	MVE7	191	579	-0.25 **	0.78 **	1.22 **			0.63	0.63	-5.0	-3.1
15	SMP4	227	579	-0.19 **	0.96 **	2.01 **			1.15	1.29	1.6	2.0
16	BOX9	261	579	-0.31 **	-0.12 **	1.44 **			1.37	1.34	3.9	1.9
17	TBL1	132	193	-1.35 **					1.55	1.50	4.0	2.6
18	TBL2	135	193	-1.42 **					1.59	1.54	4.0	2.7
19	**TBL3**	**251**	**386**	**-0.94 ****	**-0.47 ****				**1.94**	**2.06**	**6.6**	**4.4**
20	TBL4	127	193	-0.87 **					1.23	1.22	2.6	1.6
21	TBL5	210	772	-0.44 **	0.52 **	0.62 **	2.96 **		1.21	1.15	2.4	1.2
26	MVE5	226	579	0.03 **	0.68 **	3.14 **			1.30	1.25	3.6	1.7
27	TWN1	283	579	-0.19 **	0.68 **	1.53 **			1.27	1.29	3.0	2.0
28	TWN2	244	579	0.17 **	0.77 **	1.62 **			1.29	1.32	3.2	2.0

*****Output Continues****

Items 1 to 21 and 26 to 50 2000 data (Run No C2) [Grade 9]

Item Estimates (Thresholds) In input Order all on all (N = 193 L = 46 Probability Level=0.50)

	ITEM NAME	SCORE	MAXSCR	THRESHOLD/S 1	2	3	4	5	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
29	**TWN3**	**260**	**579**	**-0.09**	**0.81**	**1.87**			**1.39**	**1.47**	**4.1**	**3.1**
				**	**	**						
30	SP6	138	193	-0.87					1.11	1.12	1.3	0.9
				**								
31	SP7	65	193	0.90					0.95	0.93	-0.9	-0.5
				**								
32	SP8	61	193	1.10					0.91	0.88	-1.4	-0.8
				**								
33	SP9	31	193	2.52					1.31	1.09	1.7	0.4
				**								
34	SP10	173	579	0.34	0.85	1.24			1.12	1.18	1.3	1.1
				**	**	**						
35	SP11A	398	579	-1.41	-0.45	0.41			1.20	1.36	2.0	2.5
				**	**	**						
36	SP11B	273	579	-0.63	0.47	0.71			1.10	1.12	1.3	1.0
				**	**	**						
37	SP11C	363	579	-1.16	-0.19	0.87			1.05	1.06	0.6	0.5
				**	**	**						
38	MVE6	150	579	0.19	1.52	2.00			0.78	0.81	-2.0	-1.2
				**	**	**						
39	BT1A	170	579	0.44	0.56	3.35			0.90	0.84	-1.4	-0.6
				**	**	**						
40	BT1B	86	579	0.72	1.84	2.08			0.84	0.76	-1.1	-1.3
				**	**	**						
41	RAN3	162	579	0.22	0.83	3.85			0.91	0.88	-1.2	-0.8
				**	**	**						
42	VAR	162	579	0.34	0.85	2.42			0.87	0.87	-1.5	-0.8
				**	**	**						
43	AVG1	151	579	0.56	0.97	2.45			0.94	0.95	-0.7	-0.2
				**	**	**						
44	DRG1	41	193	1.95					1.04	0.91	0.4	-0.3
				**								
45	M4DR	78	386	1.13	1.69				0.93	0.89	-0.6	-0.4
				**	**							
46	SP2B	365	579	-3.13	-0.82	1.73			0.84	0.83	-1.6	-1.4
				**	**	**						
47	SP3B	317	579	-0.84	0.00	1.16			1.11	1.18	1.3	1.4
				**	**	**						
48	SP4B	147	193	-0.87					0.97	1.03	-0.3	0.3
				**								
49	SP5B	286	579	-0.38	-0.24	1.37			1.17	1.18	1.9	1.0
				**	**	**						

*****Output Continues****

```
Items 1 to 21 and 26 to 50 2000 data (Run No C2) [Grade 9]
-------------------------------------------------------------------------------------------------
Item Estimates (Thresholds) In input Order all on all (N = 193 L = 46 Probability Level=0.50)
-------------------------------------------------------------------------------------------------
    ITEM NAME      |SCORE MAXSCR|  THRESHOLD/S                          |  INFT  OUTFT INFT  OUTFT
                   |            |    1       2       3       4       5  |  MNSQ  MNSQ   t     t
-------------------------------------------------------------------------------------------------
50  MVE8           |   228  579 |  -0.61    0.00    0.62                |   1.03  1.03   0.3   0.3
                   |            |     .28     .29     .30                |
                   |            |                                        |
-------------------------------------------------------------------------------------------------
Mean               |            |   0.36                                 |   1.08  1.10   0.7   0.6
SD                 |            |   1.07                                 |   0.27  0.31   2.5   1.8
=================================================================================================
```

Comment: Three items out 46 (shown **bold**) have possible misfit.

```
Items 1 to 21 and 26 to 50 2000 data (Run No C2)
-----------------------------------------------------------------------------------------
Item Fit all on all (N = 193 L = 46 Probability Level=0.50)
-----------------------------------------------------------------------------------------
INFIT
 MNSQ  0.50      0.56      0.63      0.71      0.83      1.00      1.20      1.40      1.60
---------+---------+---------+---------+---------+---------+---------+---------+---------+
  1 DIE7                                 .           |          *
  2 DIE2                                 .           |    *     .
  3 SP1                        *         .           |          .
  4 TRV1                           *     .           |          .
  5 TRV2                               . *           |          .
  6 TRV3                               .     *       |          .
  7 TRV4                               .             |          .                    *
  8 TRV5                               .   *         |          .
  9 TRV6                               .         *   |          .
 10 MVE1                               .             |    *     .
 11 MVE2                               .             |  *       .
 12 MVE3                  *            .             |          .
 13 MVE4                               .            *|          .
 14 MVE7                      *        .             |          .
 15 SMP4                               .             |       *  .
 16 BOX9                               .             |          .  *
 17 TBL1                               .             |          .             *
 18 TBL2                               .             |          .               *
 19 TBL3                               .             |          .                                *
 20 TBL4                               .             |          *  .
 21 TBL5                               .             |         *   .
 26 MVE5                               .             |             *
 27 TWN1                               .             |            * .
 28 TWN2                               .             |             *
 29 TWN3                               .             |             .    *
 30 SP6                                .             |    *        .
 31 SP7                                .           * |             .
 32 SP8                                .         *   |             .
 33 SP9                                .             |             .*
 34 SP10                               .             |    *        .
 35 SP11A                              .             |       *     .
 36 SP11B                              .             |   *         .
 37 SP11C                              .             |  *          .
 38 MVE6                              .*             |             .
 39 BT1A                               .       *     |             .
 40 BT1B                               .    *        |             .
 41 RAN3                               .        *    |             .
 42 VAR                                .      *      |             .
 43 AVG1                               .          *  |             .
 44 DRG1                               .             | *           .
 45 M4DR                               .         *   |             .
 46 SP2B                               .    *        |             .
 47 SP3B                               .             |    *        .
 48 SP4B                               .            *|             .
 49 SP5B                               .             |       *     .
 50 MVE8                               .             |*            .
=========================================================================================
```

Comment: Pattern of item fit subject to constraints due to smaller sample and restricted range of achievement. *Decision*: New item (shown **bold**) has acceptable fit.