International Education Journal, 2006, 7(2), 150-159.
ISSN 1443-1475 © 2006 *Shannon Research Press*.
http://iej.cjb.net

A method for monitoring sub-trends in country-level mathematics achievement on TIMSS

Kelvin Gregory
School of Education, Flinders University *kelvin.gregory@flinders.edu.au*

The Trends in International Mathematics and Science studies provide country-level data for tracking changes in student achievement over time. In this paper the author has developed a method for identifying and monitoring trends in student achievement above or below any specified cut-point on these tests. The method involved the use of the Foster, Greer, and Thorbecke indices, as well as a modified version of these indices. The ability to identify and monitor trends in student achievement at various cut-points on the test should prove useful to policy analysts as well as to governmental and international funding agencies wishing to obtain data on the effectiveness of various programs and policies.

Monitoring trends, achievement, large-scale assessment

INTRODUCTION

Since 1995, the International Association for Evaluation of Educational Achievement (IEA) has conducted three large-scale comparative studies of mathematics and science achievement. These Trends in International Mathematics and Science Studies (TIMSS), conducted in 1995, 1999, and 2003, built on earlier IEA studies (Martin et al., 2004; Mullis et al., 2004), and involved over 50 countries. A significant proportion of these countries participated with the assistance of the World Bank and other development agencies. These funding agencies often wish to use the TIMSS data to monitor achievement and inform educational policy in the developing countries (Gilmore, 2005). More generally, participating countries are concerned with raising the level of student performance in their education systems; perhaps most of all in the case of their lowest performing students. This paper explores ways of summarising the performance of lower achieving students on TIMSS with a view to monitoring changes in such performance over time. The concepts and methods (e.g., the use of indices to monitor changes) used are drawn from the literature on poverty.

Sen (1976), as well as later researchers who picked up on his ideas, viewed poverty measurement as involving two steps: the identification of the poor and the aggregation of data on poverty into an overall index. By definition, a poor person was someone who fell below a poverty line, usually defined as an income level. The aggregation step involved the application of a rule or formula. The resulting index should be sensitive to inequality among the poor (Sen, 1976). One such group of indices was developed by Foster, Greer, and Thorbecke (1984), and is now widely known as the FGT indices. With a slight change in the basic index formulation, these indices can be easily adapted to describe mathematics performance above or below a particular cut-point on a test. The result is a new class of indices that are useful in monitoring changes in the performance of lower achieving students over time. The rest of this paper describes this new class of achievement indices, and then applies them to data from the TIMSS 1995, 1999, and 2003 mathematics assessments.

ADAPTING FOSTER, GREER, AND THORBECKE'S POVERTY INDICES

A competency cut-point is an achievement level such that students whose achievement is lower than the cut-point fail, and students whose achievement is equal to or higher than the cut-point pass. The difference between the failing student's score (θ_i) and the cut-point (z) can be defined as the score shortfall or deficit (g_i). If the student's score is equal to or above the cut-point, then the shortfall is zero by definition. The split function describing the computation of the shortfall score is:

$$g_i = \begin{cases} 0 & \theta_i \geq z \\ z - \theta_i & \theta_i < z \end{cases} \quad (1)$$

Following Foster, Greer, and Thorbecke (1984), a number of failure indices can be developed to summarise the shortfall within any population or sub population. These indices can be represented by the following formula:

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{g_i}{z}\right)^\alpha \quad (2)$$

Where q is the number of students inside the shortfall region and n is the sample size. The parameter α measures the sensitivity of the index to the degree of failure of those classified by the benchmark as having a value less than z, and usually assumes values of 0, 1, 2, and so on. This index nests several special cases. If $\alpha = 0$ the index is the proportion of students below the cut-point. If $\alpha = 1$, the index is the average of the proportionate shortfall gaps. When $\alpha = 2$, the proportionate shortfall gaps are weighted so that a doubling of the proportionate shortfall gap contributes four times as much to the index. And when $\alpha = 3$, a doubling of the proportionate shortfall gap contributes nine times as much to the index. Practically speaking then, a low index value when $\alpha = 0$ means that relatively few students are below the cut-point, while high index values, when $\alpha = 2$ or $\alpha = 3$, indicates that there are a significant number of students who have very low scores at some distance from the cut-point.

The Foster, Greer, and Thorbecke indices enable the specification of different poverty lines, consistent with the fact that such lines vary from country to country. However, if the same cut-point is used across countries, then the denominator can be removed from the indices with no loss of information. A further refinement lies in the sample divisor. In its current form, the indices are summed over q points, the number of students equal below the line, and then expressed in numerical terms with reference to the sample. That is, the indices are divided by the total sample size. One interpretation difficulty with this method is that if a sizeable proportion of the sample is at or above the cut point, the indices become relatively insensitive to changes below the cut-point. Another interpretation difficulty lies in the scale properties of the indices. The scale metric is lost when the indices are computed by dividing the shortfall by the cut-point. For these reasons, the following group of indices, called the modified FGT indices or P_β, is developed:

$$P_\beta = \frac{\sqrt[\beta]{\sum_{i=1}^{q}(g_i)^\beta}}{q}, \beta \geq 1 \quad (3)$$

In this group of indices β is an integer greater than zero. When $\beta = 1$, the index is the average distance from the cut-point for those below that point. For $\beta = 2$, the index is the average of the square root of the sum of squared shortfalls and is computationally similar to the standard deviation. When a modified index is combined with the original FGT index with $\alpha = 0$, the resulting index is expressed over the sample.

$$P = \left(\frac{\sqrt[\beta]{\sum_{i=1}^{q} (g_i)^{\beta}}}{q} \right) x \left(\frac{\sum_{i=1}^{q} (g_i)^{0}}{n} \right) = \frac{\sqrt[\beta]{\sum_{i=1}^{q} (g_i)^{\beta}}}{n}$$

or

$$P = \left(\frac{\sqrt[\beta]{\sum_{i=1}^{q} (g_i)^{\beta}}}{q} \right) x \left(\frac{q}{n} \right) = \frac{\sqrt[\beta]{\sum_{i=1}^{q} (g_i)^{\beta}}}{n} \quad (4)$$

The FGT shortfall indices are all additively decomposable. That is, each index can be decomposed to yield index values for mutually exclusive and exhaustive sub groups. For example, the indices can be decomposed to yield values for male and female students:

$P_\alpha = P_{g\alpha} + P_{b\alpha}$

In this manner, comparisons can be made of various sub groups of interest to policy makers and the like. However this property does not generally apply to the modified indices except in the special case when the sub-groups are of equal size and $\beta = 1$.

TIMSS MATHEMATICS ACHIEVEMENT DATA AND SHORTFALL INDICES

TIMSS used Bayesian population estimates that employ plausible or imputed values methods to overcome problems associated with distributing a large number of test items across several test booklets. The procedures used to obtain these Bayesian estimates for TIMSS 1995 and 1999 were described by Yamamoto and Kulick (2000) and Gonzalez, Galia, and Li (2004) for TIMSS 2003. The Bayesian population estimates were obtained by randomly drawing values from a distribution of possible values formed for each student. For both mathematics and science, five plausible were drawn for each assessed student. When calculated over all participating countries, the average of the five plausible values for mathematics would be 500 scale points, and the standard deviation would be 100 (using the original TIMSS scale). These mean and standard deviation statistics were calculated by computing the mean and standard deviation for each plausible value, and then calculating the average of these values

The shortfall indices, adapted to use all five plausible values for each student, are as follows:

$$g_{ik} = \begin{cases} 0 & \theta_{ik} \geq z \\ z - pv_{ik} & \theta_{ik} < z \end{cases} \quad (5)$$

where pv_{ik} is the kth plausible value for the ith student. The FGT-type indices are calculated by averaging over all plausible values.

$$P_\alpha = \frac{1}{5} \sum_{k=1}^{5} \frac{1}{n} \left(\sum_{i=1}^{q_k} \left(\frac{g_{ik}}{z} \right)^{\alpha} \right) \quad (6)$$

Note that the number of students falling below the cut-point can vary from plausible value to plausible value. Similar changes can be made to the modified indices to utilise the five plausible values.

Countries participating in TIMSS typically used stratified, cluster-sampling strategies (Foy, 2000). These sampling designs were considered efficient ways of obtaining representative achievement data from education systems. Typically, countries sampled intact mathematics classrooms from randomly sampled schools that were selected using a probability proportional to size method. Thus, the calculation of the indices required the use of an appropriate set of weights. In addition, the design effects associated with such sampling plans should be taken into account when calculating the standard errors of the shortfall indices. The analyses reported here use student weights and an implementation of the jackknife procedure (Gonzalez and Miles, 2001).

Cut-points are typically determined by specific educational, psychometric, or policy criteria. However, for illustrative purposes an arbitrary cut-point is chosen in this article. Since the TIMSS scales were designed to have a mean of 500, based upon a 1995 cohort, the choice of 500 as the cut-point is reasonable. This value has served as the mathematics scale reference point in the last two TIMSS assessments and was the average mathematics performance of grade 8 students participating in the 1995 assessment (Mullis et al, 2004). The analyses reported here involved the calculation of mathematics shortfall indices using $\alpha = 0$ for the FGT index and $\beta = 1$ and 2 for the modified FGT indices for those countries that participated in TIMSS 1999 and at least one of the other TIMSS assessments. In order to both simply the indices and communicate more succinctly the characteristic of each index, the following nomenclature is used:

$B_{500\alpha 0}$ - the index is referring to students below the 500 cut-point and using an alpha coefficient of zero and the original FGT formula,

$B_{500\beta 1}$ - the index is referring to students below the 500 cut-point and using the modified index with a beta coefficient of one,

$B_{500\beta 2}$ - the index is referring to students below the 500 cut-point and using the modified index with a beta coefficient of two.

Significance testing was performed using a two-tailed alpha level of 0.05, adjusted for multiple comparisons using the Bonferroni method. This was a conservative method and might serve to mask important changes at the country level.

RESULTS

When the FGT shortfall index exponent is zero, the index $\beta_{500\alpha 0}$ yields the percent of students whose achievement is below the 500 cut-point. As shown in Table 1, the index is fairly stable in some countries. For example, variations across the assessments of less than four percent are observed in England, Hungary, Republic of Korea, the Philippines, Romania, and the United States. In some countries, there is a sharp increase in the index from TIMSS 1995 to TIMSS 1999. In at least two of these cases, Israel and Italy, this increase can be explained by a change in the sampling coverage. In the case of Israel, the 1999 sample included Arab-speaking schools while the 1995 study did not. Interestingly, the percent of students in the shortfall region in Israel decreased from 1999 to 2003. For Italy, the 1999 sample represented the entire country while the 1995 sample represented only those provinces that chose to participate. Other countries with significant increases in students falling within the region from 1995 to 1999 included the Czech Republic, Iran, Singapore, and Thailand. Tunisia and Belgium (Flemish) showed significant increases from 1995 to 2003.

Table 1: Percent of students below the International Mathematics Mean (500) in TIMSS 1995, 1999, and 2003 ($\beta_{500\alpha0}$)

Country	1995		1999			2003			
Australia	38.31	(1.84)	35.79	(2.55)		47.67	(2.46)		
Belgium (Flemish)	23.68	(2.87)	20.12	(1.39)		26.57	(1.34)		⇩
Bulgaria	40.55	(2.51)	43.91	(2.78)		60.37	(2.07)	↑	↑
Canada	37.25	(1.09)	33.1	(1.01)		----	----		
Chile	----	----	89.46	(1.65)		90.35	(.82)		
Chinese Taipei	----	----	19.05	(1.10)		20.72	(1.4)		
Cyprus	59.93	(1.07)	58.61	(0.93)		66.63	(.78)	↑	↑
Czech Rep.	28.77	(1.81)	40.95	(2.46)	↑	----	----		
England	50.08	(1.50)	52.54	(2.32)		52.74	(2.95)		
Finland	----	----	36.24	(1.61)		----	----		
Hong Kong	16.87	(2.49)	13.04	(1.63)		11.85	(1.42)		
Hungary	36.38	(1.68)	34.21	(1.67)		35.21	(1.74)		
Indonesia	----	----	83.21	(1.25)		84.06	(1.31)		
Iran, Islamic Rep.	84.85	(1.22)	82.38	(1.34)		87.93	(0.74)		⇧
Israel	37.68	(2.95)	61.21	(1.72)	↑	50.65	(1.69)	↑	⇩
Italy	51.92	(1.75)	57.6	(1.83)		57.23	(1.65)		
Japan	14.85	(0.54)	16.14	(.59)		17.85	(0.73)		
Jordan	----	----	74.9	(1.37)		79.73	(1.50)		
Korea, Rep. of	16.07	(0.72)	13.66	(0.60)		13.88	(0.59)		
Latvia	54.65	(1.72)	47.41	(1.74)		44.71	(1.69)	↓	
Lithuania	61.74	(2.10)	59.32	(2.17)		48.13	(1.48)	↓	⇩
Macedonia, Rep. of	----	----	71.62	(1.54)		75.78	(1.54)		
Malaysia	----	----	41.01	(2.44)		46.4	(2.35)		
Moldova, Rep. of	----	----	64.34	(1.93)		66.47	(2.00)		
Morocco	----	----	97.54	(0.27)		95.53	(0.46)		⇩
Netherlands	33.48	(3.30)	25.88	(3.74)		30.36	(2.18)		
New Zealand	48.64	(2.31)	52.41	(2.60)		53.35	(2.68)		
Philippines	----		94.87	(0.99)		90.93	(1.30)		
Romania	58.19	(2.19)	59.9	(2.46)		59.66	(2.10)		
Russian Federation	36.66	(2.79)	37.66	(2.79)		45.58	(2.02)		
Singapore	3.65	(0.61)	9.99	(1.60)	↑	11.07	(1.34)	↑	
Slovak rep.	32.57	(1.56)	32.06	(2.04)		45.74	(1.75)	↑	⇧
Slovenia	34.11	(1.51)	35.66	(1.51)		53.84	(1.27)	↑	⇧
South Africa	94.36	(1.86)	96.16	(0.85)		95.59	(1.09)		
Thailand	40.49	(2.91)	65.87	(2.49)	↑	----	----		
Tunisia	----	----	78.86	(1.21)		92.35	(0.82)		⇧
Turkey	----	----	79.39	(1.61)		----	----		
United States	51.21	(2.38)	48.28	(1.81)		47.7	(1.77)		

↑ = significant increase from TIMSS 1995 ↓ = significant decrease from TIMSS 1995

⇧ = significant increase from TIMSS 1999 ⇩ = significant decrease from TIMSS 1999

When the shortfall exponent is 1, the modified index $\beta_{500\beta1}$ produces the average shortfall of those students below the cut-point. In Table 2 the results of these calculations are presented. The average shortfall ranges from a low of 26.01 (Singapore, 1995) to a high of 250.17 (South Africa, 2003). In general the average shortfall is remarkably stable across the years. For example, in 22 of the 36 countries that participated in two or more assessments**,** shown in Table 2, there is no significant change in the average shortfall. The average shortfall increased from 1995 to 1999 in Czech Republic, Israel, Singapore, and Thailand. In the case of Singapore, the average shortfall is almost doubled. The average shortfall in 2003 is higher than in 1995 in Singapore, Slovak Republic, and Slovenia. Compared with the 1999 average shortfall, the 2003 shortfall is higher in Cyprus, Slovak Republic, and Tunisia.

Downward trends in the average shortfall indicate upward trends in the achievement of students below the cut-point. Such changes are observed in Cyprus (1995 to 1999), Republic of Korea (1995 to 1999), Latvia (1995 to 2003), Lithuania (1995 to 2003), Morocco (1999 t 2003), and the Philippines (1999 to 2003). Both Morocco and the Philippines show substantial improvements in the average shortfall index.

When $\beta = 2$, the modified index $\beta_{500\beta2}$ provides the average of the square root of the sum of squared shortfalls. This index is more sensitive to extreme values. Thus a number of students with very low scale scores make a disproportionately high contribution to the index compared to students closer to the cut-point. The modified shortfall index ($\beta = 2$) values are presented in Table 3. The index values range from a low of 12.68 (Singapore, 1995) to a high of 705.91 (South Africa, 2003). Significant increases in the index occur from 1995 to 1999 in Czech Republic, Singapore, Slovenia, and Thailand, while a decrease is recorded in Cyprus. Compared with the 1995 index, Singapore, Slovak Republic, and Slovenia have higher index values in 2003 while Cyprus, Italy, Latvia, and Lithuania have lower values. Tunisia and Slovak Republic have higher values in 2003 compared to 1999, while Chinese Taipei, Israel, Morocco, and the Philippines have significantly lower values. Interestingly, the Moroccan 1999 value is approximately twice than of the 2003 index, indicating a substantial improvement in the lower performing students.

The shortfall indices are particularly useful in tracking changes in performance within a population. For example, Bulgaria's mean mathematics score is seen to decline from 527 in TIMSS 1995, 511 in TIMSS 1999, to 476 in TIMSS 2003 (Mullis et al, 2004). As shown in Table 1, the percentages of students falling below 500 do not change appreciably between 1995 and 1999, but do increase markedly in 2003. From the Table 2 it is suggested that much of the change in performance from 1995 to 1999 may be attributed to a decline in performance of high performing students since there is a slight, but not significant, decrease in average shortfall in 1999 compared with 1995. However, the Bulgarian average shortfall in TIMSS 2003 is substantially larger than in the earlier assessments. From the combined data in Tables 1 and 2, it is suggested that there was a dramatic and widespread decrease in Bulgarian performance on the TIMSS 2003 mathematics assessment.

DISCUSSION

In this paper a new class of indices useful in summarising changes in achievement is presented. The new indices, based upon the Foster, Greer, and Thorbecke (1984) indices, were applied to the TIMSS mathematics data. Trends in performance below the international mean of 500 are monitored, and the new class of indices appears to be useful in detecting changes in performance over time.

Table 2: Average shortfall of students below TIMSS International Mathematics Mean for TIMSS 1995, 1999, and 2003 ($\beta_{500\beta 1}$)

Country	1995		1999			2003			
Australia	67.09	(2.59)	59.07	(2.31)		64.84	(3.14)		
Belgium (Flemish)	53.97	(6.29)	51.91	(6.62)		58.50	(3.84)		
Bulgaria	67.39	(2.37)	65.99	(2.20)		77.21	(2.39)		
Canada	53.59	(1.77)	50.94	(1.31)		----	----		
Chile	----	----	124.92	(2.36)		129.62	(2.62)		
Chinese Taipei	----	----	73.74	(2.19)		62.73	(2.15)		⇩
Cyprus	92.19	(1.67)	76.37	(1.36)	↓	83.96	(1.39)	↓	⇧
Czech Rep.	43.52	(1.80)	54.12	(1.75)	↑	----	----		
England	69.23	(2.11)	66.52	(2.20)		61.64	(3.25)		
Finland	----	----	47.84	(1.63)		----	----		
Hong Kong	63.75	(6.16)	47.31	(5.82)		49.53	(6.02)		
Hungary	57.39	(2.33)	61.20	(2.12)		54.97	(2.21)		
Indonesia	----	----	127.91	(3.58)		115.17	(4.05)		
Iran, Islamic Rep.	102.79	(3.27)	103.43	(1.94)		105.89	(1.95)		
Israel	67.48	(4.39)	91.86	(3.35)	↑	71.36	(2.25)		⇩
Italy	79.91	(2.78)	78.47	(2.46)		68.65	(2.07)		
Japan	45.52	(1.34)	48.06	(1.64)		48.20	(1.52)		
Jordan	----	----	114.83	(2.12)		107.24	(2.48)		
Korea, Rep. of	57.56	(2.46)	47.15	(1.39)	↓	53.44	(1.66)		
Latvia	69.01	(2.55)	60.16	(1.83)		56.81	(1.78)	↓	
Lithuania	78.16	(2.56)	69.65	(2.76)		64.52	(1.57)	↓	
Macedonia, Rep. of	----	----	97.08	(3.06)		100.22	(2.70)		
Malaysia	----	----	58.23	(2.15)		56.82	(1.88)		
Moldova, Rep. of	----	----	80.77	(2.03)		83.13	(2.58)		
Morocco	----	----	168.66	(1.58)		119.30	(1.95)		⇩
Netherlands	56.48	(7.20)	53.53	(5.68)		46.53	(3.50)		
New Zealand	67.43	(2.63)	76.69	(2.37)		65.66	(3.58)		
Philippines	----	----	166.07	(4.38)		138.31	(3.88)		⇩
Romania	88.52	(2.90)	86.32	(3.60)		84.13	(2.85)		
Russian Federation	59.83	(2.31)	60.20	(3.00)		58.91	(1.82)		
Singapore	26.01	(1.45)	45.89	(4.06)	↑	45.78	(2.59)	↑	
Slovak rep.	52.03	(1.66)	49.80	(1.76)		63.76	(2.16)	↑	⇧
Slovenia	47.98	(1.41)	56.70	(2.07)		59.12	(1.42)	↑	
South Africa	238.59	(7.07)	236.28	(4.10)		250.17	(3.67)		
Thailand	58.07	(2.22)	80.06	(2.34)	↑	----	----		
Tunisia	----	----	74.75	(1.30)		99.75	(1.52)		⇧
Turkey	----	----	102.17	(2.20)		----	----		
United States	73.20	(2.97)	71.17	(2.01)		63.21	(1.87)		

↑ = significant increase from TIMSS 1995 ↓ = significant decrease from TIMSS 1995

⇧ = significant increase from TIMSS 1999 ⇩ = significant decrease from TIMSS 1999

Table 3: Average of the square root of squared shortfalls for students below TIMSS International Mathematics Mean for TIMSS 1995, 1999, and 2003 ($\beta_{500\beta 2}$)

Country	1995		1999			2003			
Australia	74.54	(5.31)	56.90	(3.97)		66.88	(6.49)		
Belgium (Flemish)	53.01	(12.20)	47.04	(14.00)		59.32	(7.69)		
Bulgaria	72.10	(4.29)	69.91	(3.86)		91.85	(5.30)		
Canada	47.86	(3.01)	43.02	(1.99)		----	----		
Chile	----	----	205.75	(6.58)		214.72	(7.41)		
Chinese Taipei	----	----	92.21	(5.49)		64.09	(4.12)		⇩
Cyprus	132.61	(4.61)	91.34	(2.88)	↓	106.11	(3.35)	↓	
Czech Rep.	31.07	(2.49)	48.02	(2.93)	↑	----	----		
England	76.48	(4.25)	71.11	(4.39)		58.14	(5.26)		
Finland	----	----	39.37	(2.78)		----	----		
Hong Kong	71.66	(12.78)	42.62	(12.45)		43.92	(9.71)		
Hungary	54.01	(4.23)	63.02	(4.59)		50.39	(4.43)		
Indonesia	----	----	225.55	(10.48)		182.27	(11.84)		
Iran, Islamic Rep.	146.44	(8.94)	149.97	(4.77)		147.92	(4.74)		
Israel	82.78	(10.79)	131.08	(8.84)		79.19	(4.55)		⇩
Italy	104.85	(6.45)	96.34	(5.56)		73.09	(4.24)	↓	
Japan	36.25	(2.26)	41.30	(2.95)		39.99	(2.42)		
Jordan	----	----	190.82	(6.37)		163.04	(6.61)		
Korea, Rep. of	58.19	(5.15)	39.28	(2.63)		49.08	(2.99)		
Latvia	73.76	(5.51)	58.23	(3.71)		50.71	(2.80)	↓	
Lithuania	94.62	(5.67)	73.73	(5.45)		64.48	(3.02)	↓	
Macedonia, Rep. of	----	----	142.65	(7.62)		147.56	(7.45)		
Malaysia	----	----	55.26	(3.98)		48.86	(2.97)		
Moldova, Rep. of	----	----	97.05	(4.6)		104.18	(5.73)		
Morocco	----	----	358.53	(6.03)		181.09	(5.11)		⇩
Netherlands	58.13	(14.96)	49.15	(9.57)		35.91	(5.35)		
New Zealand	72.74	(5.19)	90.69	(4.89)		66.51	(7.47)		
Philippines	----	----	352.76	(14.84)		245.11	(11.19)		⇩
Romania	122.21	(7.33)	117.06	(8.56)		108.3	(6.40)		
Russian Federation	58.19	(3.90)	61.54	(5.53)		55.3	(3.11)		
Singapore	12.68	(1.38)	35.98	(5.60)	↑	33.74	(3.29)	↑	
Slovak rep.	46.34	(2.99)	42.26	(2.70)		65.53	(4.34)	↑	⇧
Slovenia	36.59	(1.94)	54.22	(3.61)	↑	54.4	(2.58)	↑	
South Africa	644.64	(27.77)	651.63	(16.43)		705.91	(15.36)		
Thailand	53.85	(3.87)	96.89	(4.93)	↑	----	----		
Tunisia	----	----	80.02	(2.34)		125.61	(3.28)		⇧
Turkey	----	----	145.98	(5.39)		----	----		
United States	86.76	(6.34)	79.02	(3.62)		62.36	(3.47)		

↑ = significant increase from TIMSS 1995 ↓ = significant decrease from TIMSS 1995

⇧ = significant increase from TIMSS 1999 ⇩ = significant decrease from TIMSS 1999

It is relatively easy to modify both classes of indices to monitor high performance. For example, if the desire were to track changes in performance above 600 scale points, then the split function would be:

$$g_{ik} = \begin{cases} pv_{ik} - 600 & \theta_{ik} > 600 \\ 0 & \theta_{ik} \leq 600 \end{cases} . \qquad (7)$$

Reasons for changes in index cut-point values are best provided at the local level. For example, a country may wish to monitor proficiency changes in the advanced benchmarking region of a national assessment. Nevertheless, given that the TIMSS assessments are psychometrically sound, the indices used in this article appear to be useful for monitoring changes in low performance over time. The FGT index ($\alpha = 0$) captures the percentage of students within the designated region while the modified indices provide useful summarisations of the achievement data within the region.

Issues of multidimensionality of failure arise because individuals, educators, and policy makers often need to describe achievement on several individual attributes, including knowledge, problem solving, and literacy. Multidimensional failure indices can be developed that take into account the different facets of achievement. For example, the TIMSS mathematics curriculum and assessment frameworks (Robitaille et al, 1993; Mullis et al., 2003) include a number of content areas and processes. A multidimensional mathematics failure index can include dimensions for each content and process area, and can be extended to include opportunity to learn and other factors that are shown to be related to mathematics achievement. Such an approach minimises the temptation to place undue emphasis upon an overall achievement score, and yields a richer understanding likely to inform better and more direct policy decisions. The results presented in this paper can be easily conceptualised as being weighted indices of multidimensional component indices.

REFERENCES

Foster, J., Greer, J. and Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica*, 52(3), 761-766.

Foy, P. (2000). Sampling weights. In M.O. Martin, K.D. Gregory and S.E. Stemler (Eds.), *TIMSS 1999 Technical Report* (pp. 189-202). Boston, MA: International Study Center, Boston College.

Gonzalez, E.J. and Miles, J.A. (2001). *TIMSS 1999 User Guide for the International Database*. Boston, MA: International Study Center, Boston College.

Gonzalez, E.J., Galia, J. and Li, I. (2004). Scaling methods and procedures for the TIMSS 2003 mathematics and science scales. In. M.O., I.V.S. Mullis and S.J. Chrostowski, (Eds.), *TIMSS 2003 Technical Report*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Gilmore, A. (2005). The impact of PIRLS (2001) and TIMSS (2003) in low-and middle-income countries: An evaluation of the value of World Bank support for international surveys of reading literacy (PIRLS) and mathematics and science (TIMSS). [Online] http://www.iea.nl/iea/hq/fileadmin/user_upload/WB-report.pdf [June 26, 2005].

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J. and Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J. and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J. and O'Connor, K.M. (2003). *TIMSS Assessment Framework and Specifications.* (Second Ed). Chestnut Hill, MA.: Boston College..

Robitaille, D.F., McKnight, C.C., Schmidt, W.H., Britton, E., Raizen, S. and Nicol, C. (1993). *TIMSS Monograph No. 1: Curriculum Frameworks for Mathematics and Science*. Vancouver: Pacific Educational Press.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, 44(2), 219-231.

Yamamoto, K. and Kulick, E. (2000). Scaling methodology and procedures for the TIMSS mathematics and science scales. In M.O. Martin, K.D. Gregory, and S.E. Stemler (Eds.), *TIMSS 1999 Technical Report* (pp. 237-263). Boston, MA: International Study Center, Boston College.