This PowerPoint was used in a presentation to Hong Kong Education Commission's 2006 Reporting Session on Progress of Education Reform held in Hong Kong on 2 December 2006.

Barry McGaw is half-time Professorial Fellow and Director of the Melbourne Education Research Institute at The University of Melbourne and works half-time as a consultant through McGaw Group Pty Ltd.

Professor McGaw returned to Australia at the end of 2005 from Paris where he had been Director for Education at the Organisation for Economic Co-operation and Development (OECD). He had previously been Executive Director of the Australian Council for Educational Research (ACER) from 1985 to 1998 and Professor of Education at Murdoch University in Perth Western Australia from 1976 to 1984.

Professor McGaw holds a PhD in educational psychology and psychometrics from the University of Illinois and a BSc in chemistry, psychology and statistics and a BEd Hons (First Class) in educational psychology from the University of Queensland.

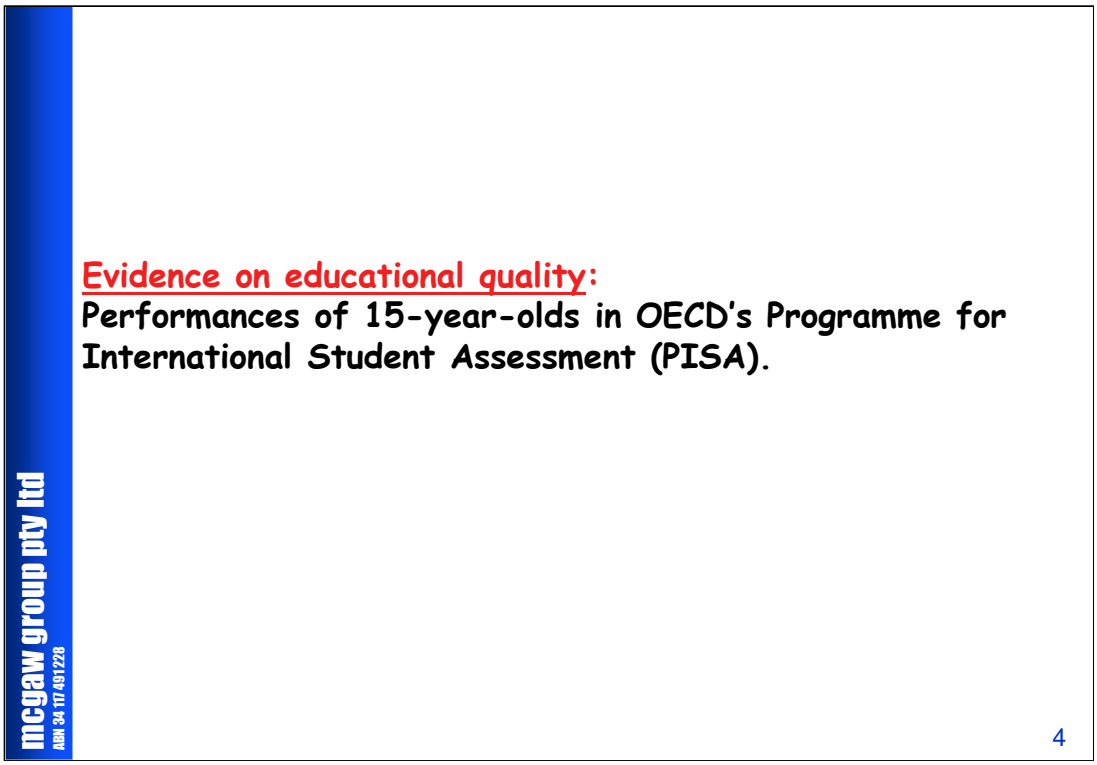
He is a Fellow of the Academy of the Social Sciences in Australia, the Australian Psychological Society and the Australian College of Educators. He has been President of the Australian Association for Research in Education, the Australian Psychological Society, the Australian College of Educators and the International Association for Educational Assessment.

Outline

- International evidence on Hong Kong education
  - Evidence on quality
  - Evidence on equity
- Curriculum reform
- Examinations
- Post-secondary education


First, the international comparative evidence…
mcgaw group pty ltd
ABN 34 117 491228

.



The evidence on the quality of the outcomes of education systems is drawn from the OECD's Programme for International Student Assessment (PISA) for which details are available on www.pisa.oecd.org.

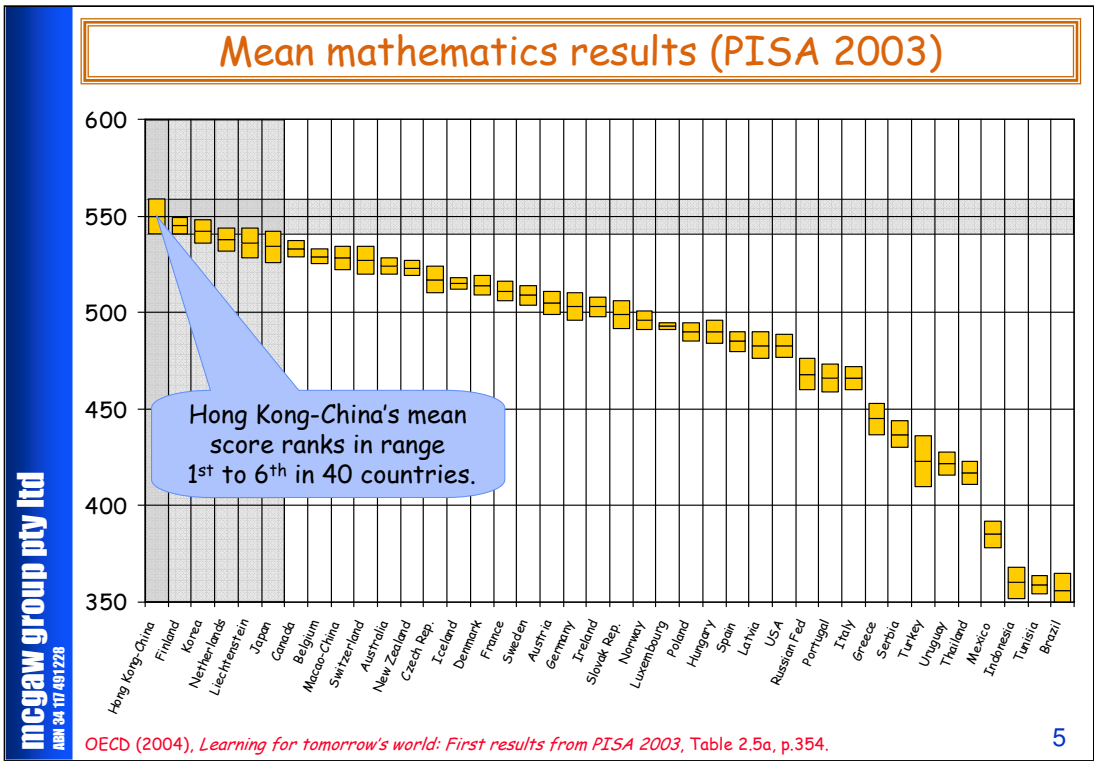
PISA provides direct, internationally comparable evidence of the quality of national education systems with its assessments of the achievements of 15-year-olds. The population assessed is 15-year-olds in schools of any type so it excludes 15-year-olds who are not in school.

In PISA 2000, students were assessed in reading literacy, mathematics and science, with reading literacy as the main domain and mathematics and science as minor domains. In PISA 2003, mathematics was the main domain and reading and science minor domains together with problem solving which was an additional domain. In PISA 2006, the three original domains have been assessed, with science as the main domain.

PISA does not assess whether students have learned the specific content of their curricula but rather whether they are able to use the knowledge and skills they have acquired. Both open-ended and multiple-choice questions are used. In the PISA 2003 mathematics assessments, for example, there were 85 items, 17 of them simple multiple choice, 11 complex multiple choice and 57 items that required students to construct their response. Sample items, illustrating the content and form of assessment, are provided on the PISA website, given above.

All assessment tasks are provided in both English and French and countries using other languages are required to produce two independent translations into their own language(s), one from the English and one from the French, and then to compare them in producing their final draft which is then independently checked by an external translator.

All potential assessment materials are first reviewed in all participating countries for prima facie evidence of cultural bias, with doubtful items being removed. All material that survives is then used in an internationally controlled trial in all participating countries a year before the actual PISA assessment. The performances of students on the trial material provide empirical evidence on whether tasks work consistently in all countries. Tasks that do not are removed from the pool of tasks from which those to be used in the final tests are selected.



When thinking about what mathematics might mean for individuals, one must consider both the extent to which they possess mathematical knowledge and understanding, and the extent to which they can activate their mathematical competencies to solve problems they encounter in life. PISA, therefore, presents students with problems mainly set in real-world situations. These are crafted in such a way that aspects of mathematics would be of genuine benefit in solving the problem. The objective of the PISA assessment is to obtain measures of the extent to which students presented with these problems can activate their mathematical knowledge and competencies to solve such problems successfully.

The figure shows the results in PISA 2003 in mathematical literacy, which was the major domain of assessment in PISA that year.

The line in the middle of each box gives the mean performance of 15-year-olds in the education system. The results reveal marked variations in performance levels among the 40 that participated – ranging from Hong Kong at the top to Indonesia, Tunisia and Brazil at the bottom.

The size of a box reflects the precision with which a country's mean is estimated, the least precise in PISA 2003 being that for Turkey. Where the boxes overlap on the vertical dimension, there is no significant difference between the means for the countries. (Further details are given in the PISA report, as indicated in the source information at the foot of the figure.)

While Hong Kong ranked in 1st place, its mean is not significantly different from those of Finland, South Korea, the Netherlands, Liechtenstein and Japan. It is, therefore, appropriate to say that Hong Kong ranked between 1st and 6th or that Hong Kong tied in 1st place with five others.

Hong Kong's performances in OECD PISA

	Reading PISA 2000	Mathematics PISA 2003	Problem solving PISA 2003
Behind	Finland	-	-
Rank	2nd to 10th	1st to 6th	1st to 6th
Tied with	Canada New Zealand Australia Ireland Hong Kong-China Korea United Kingdom Japan Sweden	Hong Kong-China Finland Korea Liechtenstein Netherlands Japan	Korea Hong Kong-China Finland Japan New Zealand Macao-China

Sources: OECD (2001) *Knowledge and skills for life: First results from PISA 2000*, Fig. 2.4, p.53.
OECD (2004), *Learning for tomorrow's world: First results from PISA 2003*, Fig 2.16b, p.92.
OECD (2004), *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003*, Fig 2.4, p.42.

6

mcgaw group pty ltd
ABN 34 117 491228

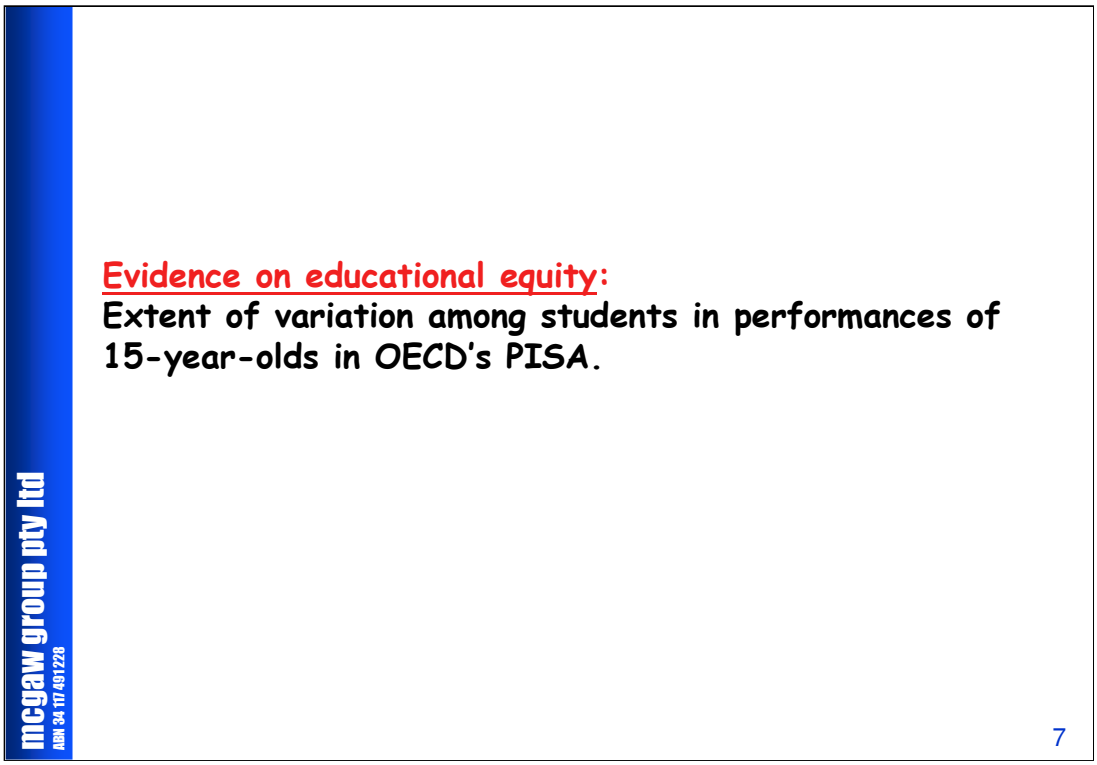
A broader picture of the average quality of the educational performances of 15-year-olds in Hong Kong can be obtained by considering the results from various PISA assessments:

In PISA 2000, when reading literacy was the main domain of assessment, Hong Kong ranked in 6th place but it was significantly behind only Finland which ranked in 1st place. Hong Kong's mean performance was not significantly different from those of 8 other countries – Canada, New Zealand, Australia Ireland, South Korea, the United Kingdom, Japan and Sweden. One can say, therefore, that Hong Kong ranked from 2nd and 10th or tied in 2nd place with 8 other countries.

In mathematics in PISA 2003, as shown in the previous slide, Hong Kong ranked 1st but was not significantly better than 5 other countries – Finland, South Korea, Liechtenstein, the Netherlands and Japan. As noted on the previous slide, Hong Kong can be said to have ranked from 1st to 6th or to have tied in 1st place with 5 other countries.

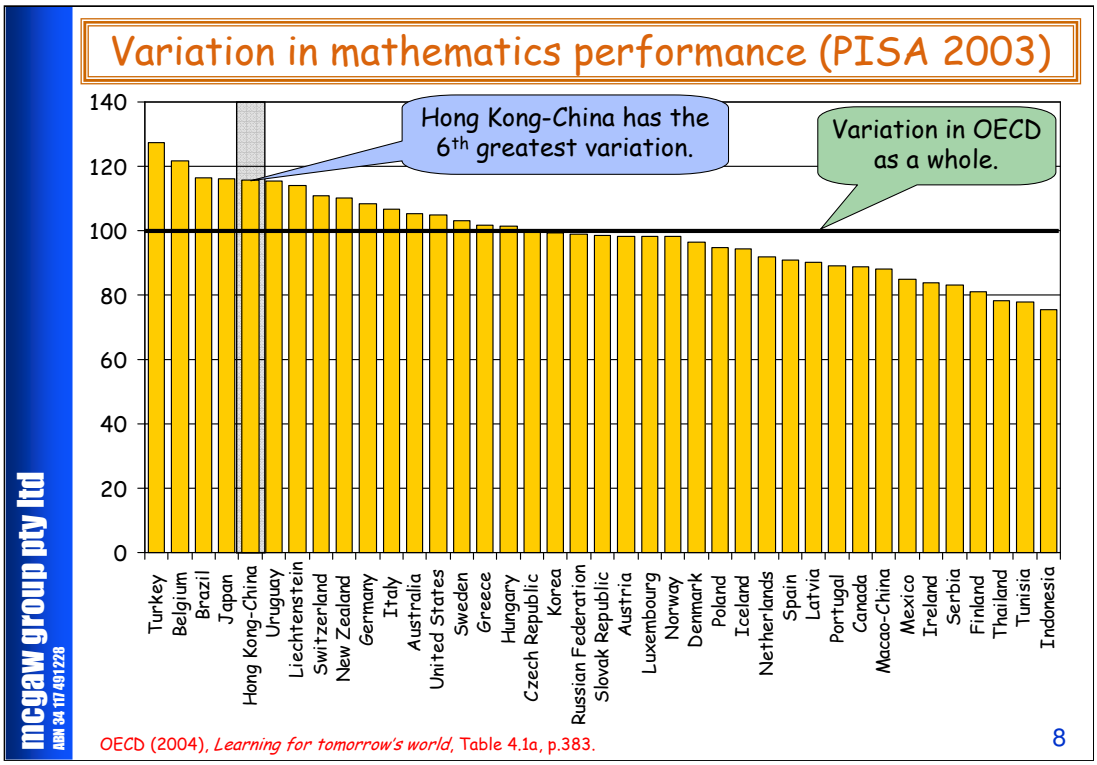
PISA 2003 also assessed problem solving. The tasks for this assessment required students to resolve real-life problems where the solution path is not immediately obvious and where the content that might be applicable is not within a single subject area. Students could not simply use some defined process that they had studied. They had to merge knowledge and strategies to resolve the problems.

In problem solving, Hong Kong ranked in 2nd place behind South Korea but not significantly different from it or from 4 others ranked behind it – Finland, Japan, New Zealand and Macao-China. n , Australia ranked in 4th place behind Korea, Finland and Japan and tied with New Zealand, Canada and Belgium. In problem solving, therefore, Hong Kong can also be said to have ranked from 1st to 6th or to have tied in 1st place with 5 others.



Quality is only one issue in evaluating education systems. Equity is also important. Perhaps the simplest way in which to think about equity is to see how far the more poorly performing students fall behind the best performers.

This is not to suggest that there should be no differences. It is only to ask what international comparisons might reveal to be possible.



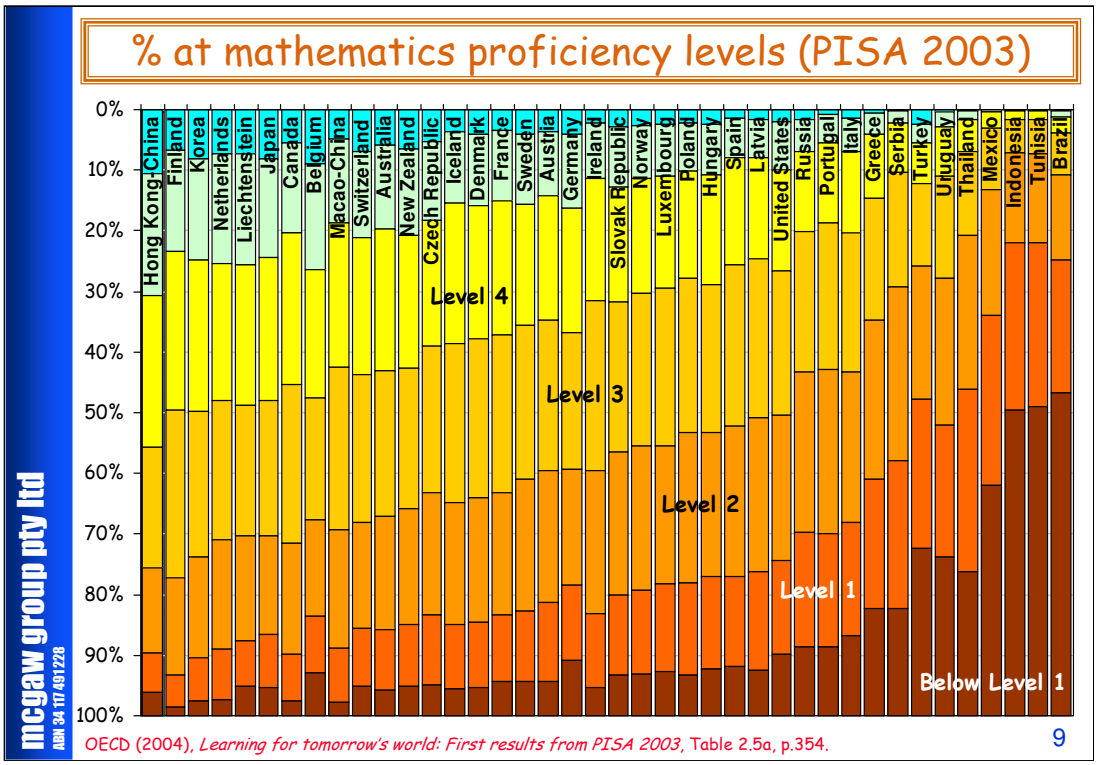
Average performance tells only part of the story on the quality of education. The variation in achievements tells another part. The current high-profile, US education act, No Child Left Behind, captures the concern that education systems might serve their best performers well but then leave many others behind.

There will always be differences in performances among students but it is interesting to compare countries in the magnitude of the variation. The figure above shows the variation for each system in mathematics in PISA 2003 on a scale on which the variation for OECD countries as a whole is set at 100.

The variation in mathematics performances in Hong Kong is 113, well above the OECD average of 100. Hong Kong has the 6th greatest variation among the 40 countries.

Some of the countries with the smallest variation stand out for two reasons. First, not all of their 15-year-olds are still in school so many of the poorer performers have left. PISA does not assess them. Secondly, these countries have very few high performers so the variation in their scores is, to some extent, reduced by having a low top end.

Hong Kong has a high average result and a relatively large proportion of students at the top end but so too do some other countries which have much smaller variation than Hong Kong, e.g. Finland for which the variation is only 81 and the Netherlands for which the variation is 94.



OECD (2004), *Learning for tomorrow's world: First results from PISA 2003*, Table 2.5a, p.354.

The nature of the variation in mathematics performances is shown in more detail in the figure above. The countries are arranged in order of their mean performance, as shown in Slide 5. As the highest in mean performance, Hong Kong is on the left.

Hong Kong has the highest proportion of students performing at Level 6, the highest level of performance and it clearly has the highest proportion at Levels 5 and 6 together. Finland and South Korea have smaller proportions at both of these levels but end up with average performances not different from Hong Kong's because they have smaller proportions of their students at the lowest levels (Level 1 and Below Level 1). Finland and South Korea produce a smaller spread of performances than Hong Kong by doing a little better with their lowest performing students and a little worse with their better performing students than Hong Kong does.

Belgium is a country where the spread is marked – second greatest as shown in the previous slide. Given where it stands on average, Belgium has a surprising high proportion of its students performing at Levels 6 and 5 but also a surprisingly large proportion of its students performing at Level 1 or Below Level 1.

In comparison with countries of the right of the figure, Hong Kong has very few students at the lowest level.

All of the results shown so far make clear that Hong Kong has a high quality education system.

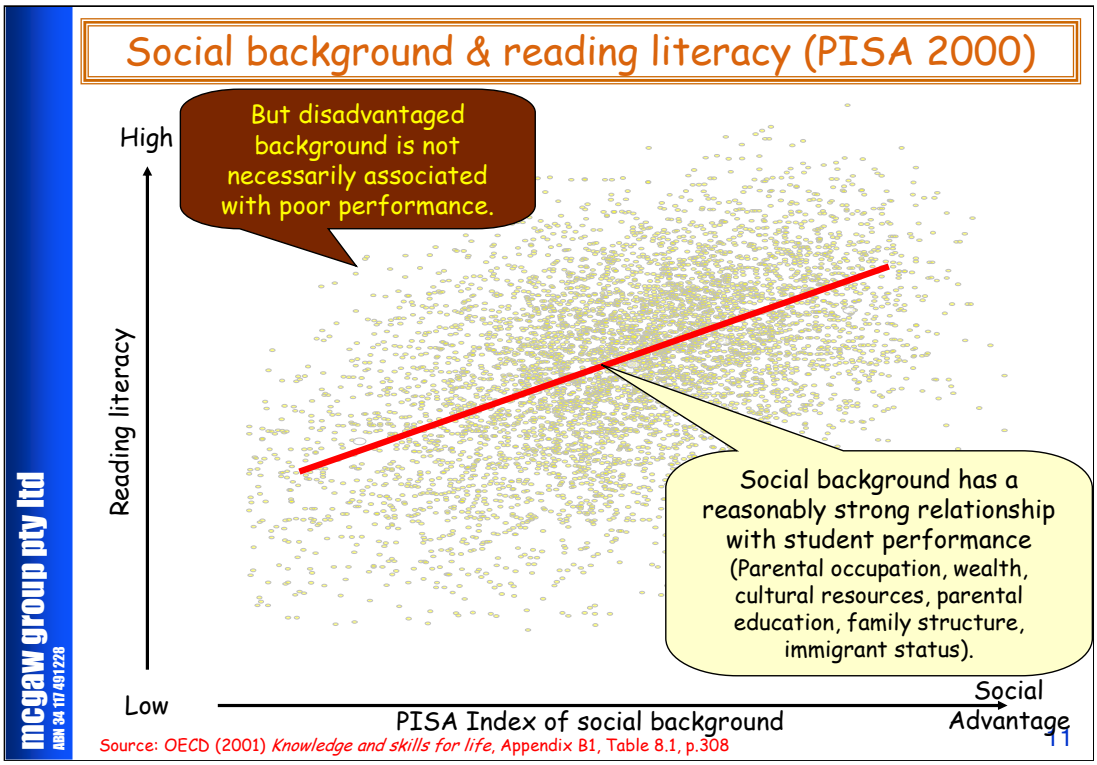
mcgaw group pty ltd
ABN 34 117 491228

Evidence on educational equity:

Relationship between the social backgrounds of 15-year-olds and their performances in OECD's PISA.

10

A more important way in which to consider equity is to examine the relationship between students' social backgrounds and their educational performance.



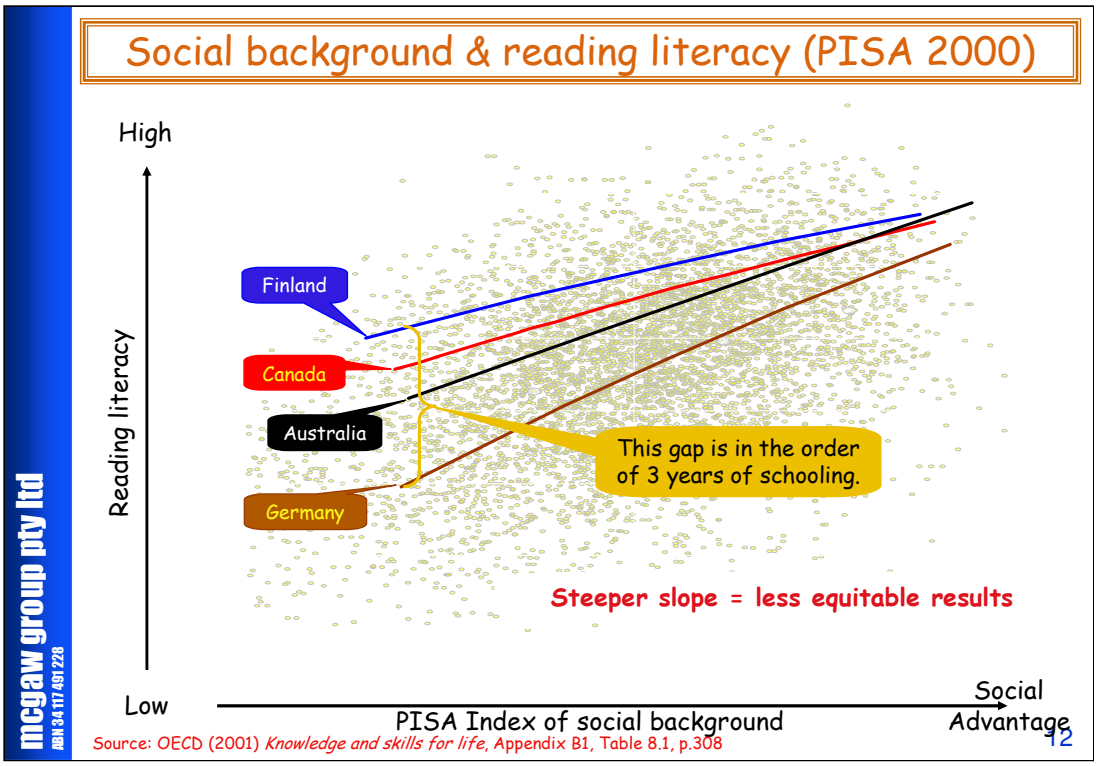
The 15-year-olds involved in PISA complete a questionnaire that collects information important for the interpretation and analysis of the results. Students are asked about characteristics, such as gender, economic and social background, and activities at home and school.

The information on economic and social background – parents' education and occupation, cultural artefacts in the home – permit the construction of an index of social background that ranges from socially disadvantaged to socially advantaged. This scale is comparable across countries.

The relationship between social background and reading literacy in PISA 2000 is shown in the figure above in which the results of the 265,000 15-year-olds in the sample on both variables are plotted. The correlation is relatively high (around 0.45) indicating quite a strong relationship between the two variables. The slope of the regression line that summarises the relationship is quite steep, indicating that increased social advantage, in general, pays off with considerable increase in educational performance.

It can, nevertheless, be seen that there are many exceptions – socially advantaged individuals who do not perform well (towards the bottom-right of the graph) and students from disadvantaged backgrounds who perform well (towards the top-left of the graph).

This result has been long established in research in many individual countries and it can lead to a counsel of despair. If the relationship between social background and educational achievement is so strong, education can seem to be impotent, unable to make a difference. There is other research evidence that provides assurance that schools can make a difference to the life chances of their students but the PISA also provide additional insights because it is possible to compare regressions lines of the type above for individual countries.

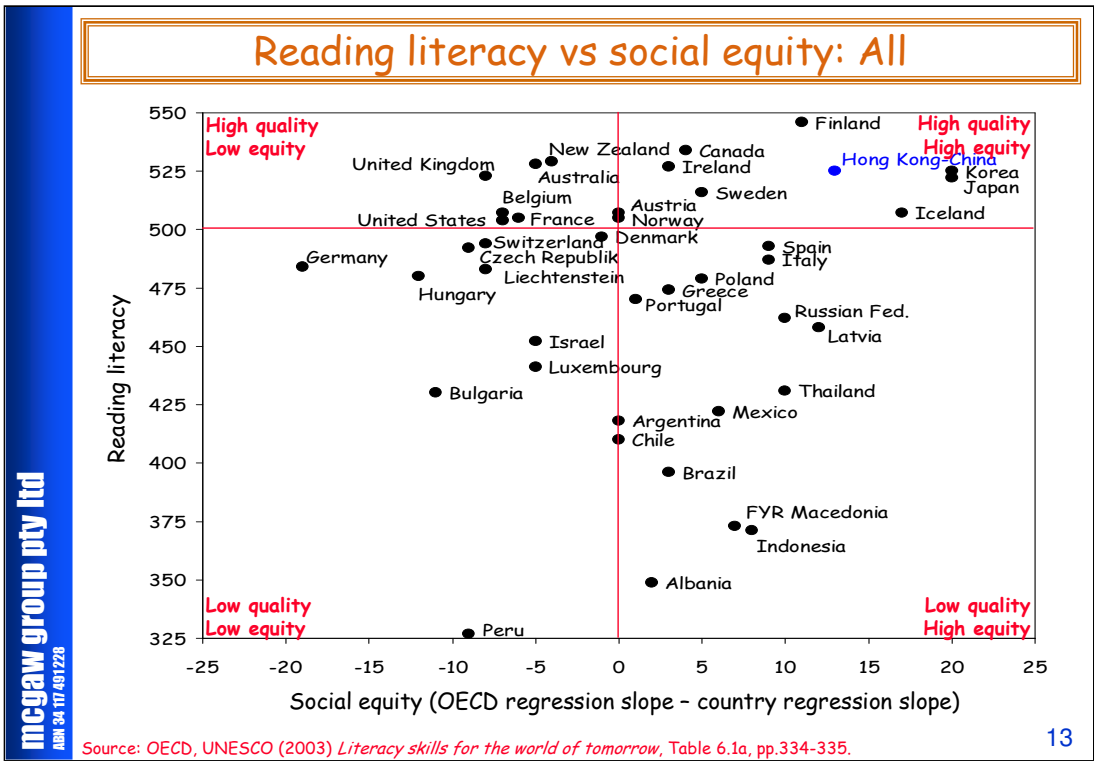


An examination of the relationship between social background and educational achievement country-by-country reveals marked differences among countries. The figure above shows the results for four countries. The lines for Finland and Canada are significantly less steep than the one for the OECD as a whole which was shown in the previous slide. Increased social advantage in these countries is associated with less increase in educational achievement than in the OECD as a whole. The results in these countries are more equitable than those of the OECD overall. Students differ in achievement but not in a way that is so substantially related to their social background.

The lines for Australia and Germany are both significantly steeper than the one for the OECD as a whole, as are those for the US and the UK which are not shown in the figure above. In all of these countries, social background is more substantially related to educational achievement than in the OECD as a whole. Their results are inequitable in the sense that differences among students in their literacy levels reflect to a marked extent differences in their social background.

The differences between these four lines at the left-hand end are substantial. Socially disadvantaged students do very much worse in some of these countries. The gap in educational achievement between similarly socially disadvantaged students in Germany and Finland represents around three years of schooling. Similarly disadvantaged students in Australia fall about half-way between, around 1½ behind their counterparts in Finland.

More detailed analysis of the German data shows the pattern to be strongly related to the organisation of schooling. From age 11, students are separated into vocational and academic schools of various types on the basis of the educational future judged to be most appropriate for them. Students from socially disadvantaged backgrounds generally end up in low-status vocational school and achieve poor educational results. Students from socially advantaged backgrounds are directed to high-status academic schools where they achieve high-quality results. The schooling system largely reproduces the existing social arrangements, conferring privilege where it already exists and denying it where it does not.



If lines for more countries were to be added to the figure on the previous slide, the pattern would become difficult to discern. The figure above provides a clearer picture in which the locations and slopes of the lines for all OECD countries are represented.

Mean performances of countries in reading literacy are represented on the vertical axis. The slope of the regression line for social background on reading literacy is represented on the horizontal axis as the difference between the slope for the OECD as a whole and a country's own slope. This places to the left countries where the slope is steeper than in the OECD as a whole (that is, countries in which social background is most substantially related to educational achievement) and to the right countries where the slope is less steep than that for the OECD as a whole (that is, countries in which social background is least related to educational achievement).

Countries high on the page are high-quality and those to the far right are high-equity. The graph is divided into four quadrants on the basis of the OECD average on the two measures.

The presence of countries in the 'high-quality, high-equity' quadrant (top right) demonstrates that there is no necessary trade off between quality and equity. They show that it is possible to achieve both together. Korea, Japan, Hong Kong, Finland and Canada are among them.

Those in the top-left quadrant are 'high-quality, low-equity', with average performances above the OECD as a whole but with relatively steep regression lines.

The United States is only average quality but it is low-equity. Germany, as a low-quality, low-equity country, is in the bottom-left quadrant along with a number of other countries that also begin to separate students into schools of different types as early as age 11-12.

Some of the countries in the 'low-quality, high-equity' quadrant achieve the 'high-equity' status only because social background differences are not strongly related to performance differences among those 15-year-olds who remain in school. The lack of equity in these systems is evident in a high incidence of early departures among those who are socially disadvantaged. In Mexico, for example, only 60% of 15-year-olds are in school.

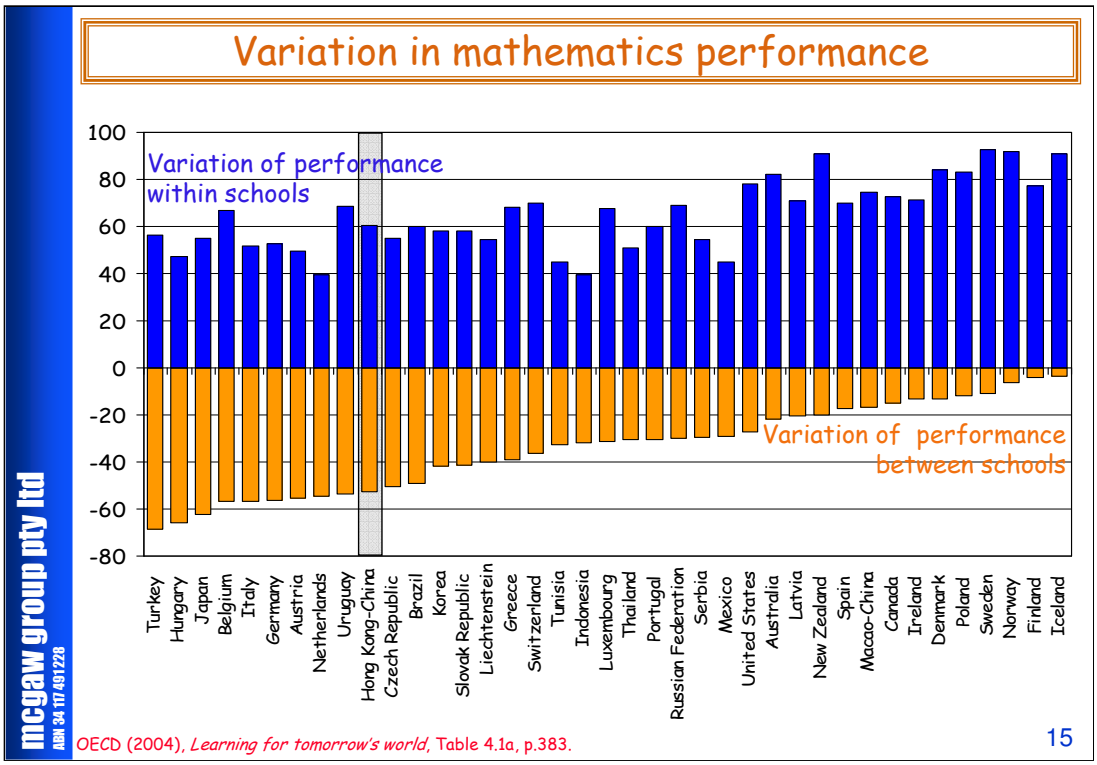
mcgaw group pty ltd
ABN 34 117 491228

Evidence on educational equity:

Sources of variation among students and schools in performances of 15-year-olds in OECD's PISA.

14

A further way in which to consider equity is to examine the magnitude of differences in performances among schools and the relationship of those differences to the social backgrounds of students.

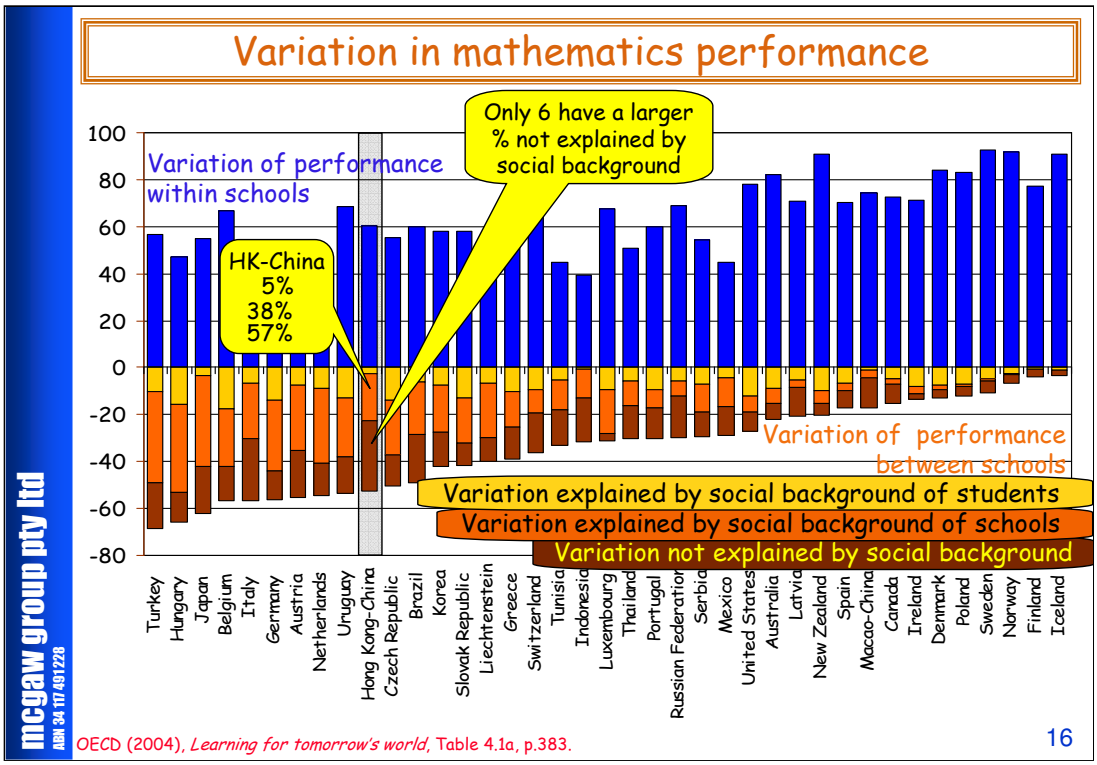


The figure above divides the variation in student performance in mathematics in PISA 2003 for each country into a component due to differences among students within schools, shown above the zero line, and a component due to differences between schools shown below that line. (The columns for each country are those in Slide 7, but with countries ordered here from left to right in terms of the magnitude of the variation between schools.

In Iceland, Finland and Norway there is very little variation in scores between schools. For parents in these countries, choice of school is not very important because there is so little difference among schools.

Among the countries in which a large percentage of the variation is between schools, there are some in which this occurs by design. In Hungary (58%) and Germany (52%), for example, students are sorted into schools of different types according to their school performance as early as age 12. This then maximises the variation between schools. In Hong Kong, 47% of the variation is between schools, in Iceland it is 4%, in Finland 5% and Norway 7%.

For Poland, in PISA 2000, 63 per cent of the variation in reading was between-schools whereas in PISA 2003 in mathematics only 13 per cent was between schools. This remarkable difference was due to a reform in which early streaming of students into schools of different types was abandoned in favour of comprehensive schools for students up to the age at which PISA measures their performance. (Not only was the between-school variation reduced. Poland was the only country to improve its average performance significantly on all measures used in both PISA 2000 and PISA 2003. It did so largely by raising the achievement levels of its poorer performing students.)

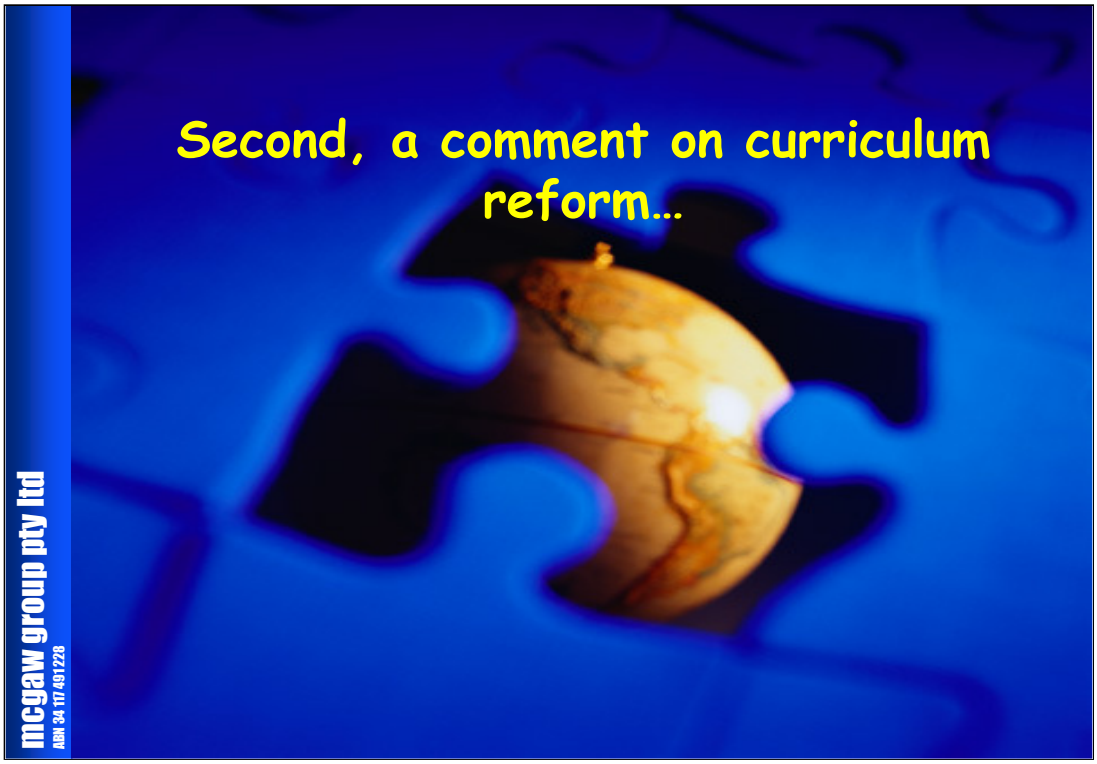


A further way in which to examine equity is to determine the extent to which the variation between schools can be explained in terms of differences in the social backgrounds of the students. This is done in the figure above, with the between-school variation subdivided into three components: (a) variation that can be accounted for in terms of the social backgrounds of the individual students in the schools (the influence of students' own social backgrounds); (b) variation that can be accounted for in terms of the average social background of the students in the schools (the influence of the company that students keep); and (c) variation that cannot be accounted for in terms of the social backgrounds of the students.

Analyses of the PISA 2000 data for Austria shows that the influence of company is not uniform. Socially advantaged students do not gain much benefit from being in the company of other advantaged students but socially disadvantaged students suffer a substantial additional disadvantage from being in the company of other socially disadvantaged students. Reducing stratification would bring benefits to the disadvantaged and no costs to the advantaged (Schneeweis & Winter-Ebmer, Peer effects in Austrian schools. Working Paper No. 0502, Department of Economics, Johannes Kepler University of Linz, Austria 2005, p.2).

In Hong Kong, 43% of the variation between-schools can be accounted for in terms of differences between schools in the social background of their students – 5% individual social background and 38% per cent the average social background of students in the schools. Only six have a smaller percentage of the between school difference that can be explained in terms of social background which makes the Hong Kong system among the most equitable by this measure.

The fact that most of the effect of social background in Hong Kong occurs at the school level and not the individual level suggests that reform of Hong Kong's School Places Allocation System will make the Hong Kong education system even more equitable.

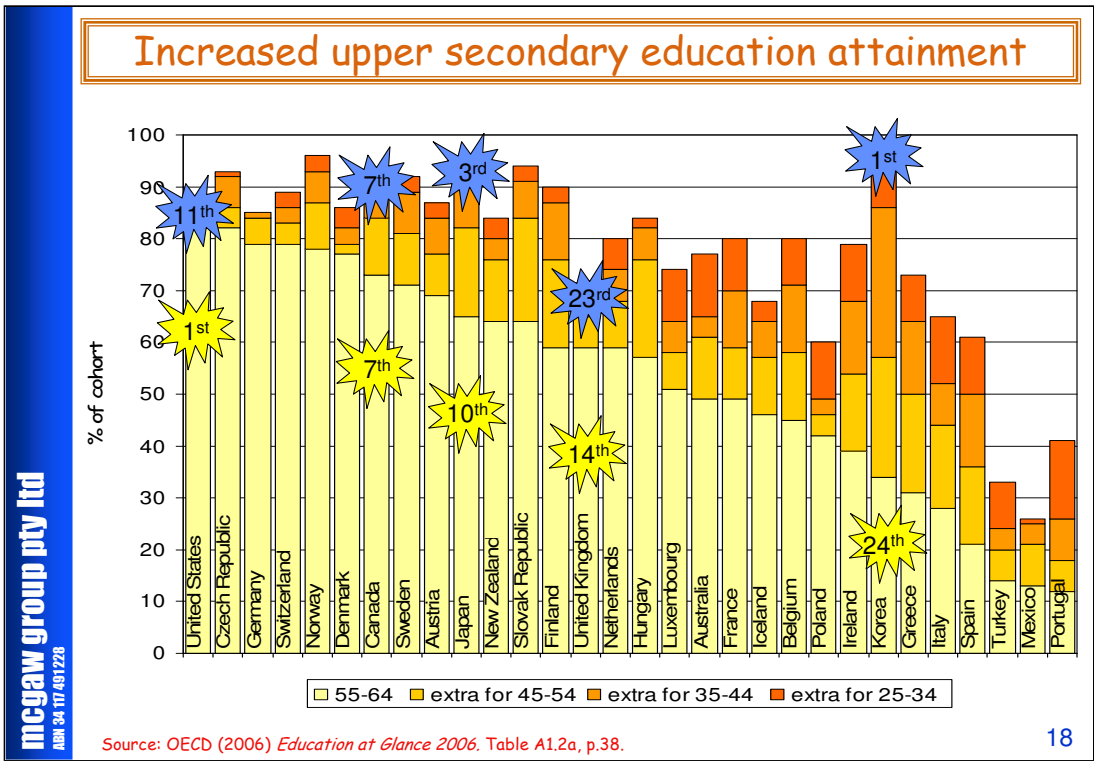


The available international comparisons of education systems provide good grounds for considerable satisfaction in Hong Kong. The Hong Kong system, at least in the middle secondary years, is high-quality and, by most measures, high-equity. That does not mean that Hong Kong has no need to change or that it could not do better.

The environment is changing in ways that put new demands on education systems. The changing economic environment places new pressures on education systems to prepare students with different and higher level knowledge and skills.

The competition is also changing. Countries shown by international comparisons, such as those of PISA, to be doing less well are committing effort and resources to catching up. Those who are ahead could fall behind if they stand still.

It is perfectly appropriate, therefore, that Hong Kong has undertaken a significant review of its system and is in the midst of reform.



One response to the new knowledge economy is that many education systems are attracting more young people to stay on to reach higher levels of initial education.

There are no internationally comparable data on trends in completion rates for upper secondary education but a picture for past decades can be obtained from the percentages of the population in different age brackets that have attained this level. This picture is available, however, only for OECD countries but it is still worth looking at and reflecting on from a Hong Kong perspective and examining with any relevant data that Hong Kong's might have available.

For OECD countries, the percentage of 55-64 year-olds who have attained upper secondary education indicates completion rates 37-46 years ago. The picture is only approximate because some will have attained this level as adults, long after having left initial education, and also because some of the population will not have survived to this age-group. Younger groups provide corresponding pictures for more recent decades.

The figure above shows the attainment rates for 55-64 year-olds in OECD countries and, for successively younger age groups, the increase in the rate compared with the next oldest group. The rates for 25-34 year-olds reveal that, by 7-16 years ago, 17 of the 30 OECD countries had achieved attainment rates of 80% or higher. Australia was not among them.

The Republic of South Korea started from a low base but grew quickly, rising from 24th to 1st. Over the same period, Japan rose from 10th to 3rd. The US started from a high base but grew quite slowly, slipping from 1st to 11th. Australian rates have grown relatively slowly from a comparatively low base, with the rank slipping marginally from equal 18th to 20th. Meanwhile Canada held its ranking at 7th.

In the mid-1960s, South Korea had a GDP per capita equivalent to that of Afghanistan and behind all the countries of Latin America. South Korea is now a Member of the OECD, with a GDP per capita that just below the top two thirds of the Members. Education reform and a deep national commitment to education and skill development are recognised as key drivers of this remarkable economic growth.

Changing structures

- Making schools more comprehensive
  - PISA evidence that differentiation exacerbates differences
  - Impact of changing
    - Polish experience with comprehensive schools raising the 'tail'
    - Will reform of Hong Kong's School Places Allocation System do the same?
    - Could Hong Kong become even higher quality and even more equitable?
- Restructuring school years
  - Hong Kong's change from 6+5+2+3 to 6+3+3+4
    - 12 years of primary and secondary education clearly the norm
    - Hong Kong and Ontario as models for coping with double cohorts finishing
  - Benefits
    - Abandoning A Levels a great move – (if only England could)
    - Broader baccalaureate structure increasingly common
    - Fits with the Bologna model (though first tertiary cycle can be 3 or 4 yrs

mcgaw group pty ltd
ABN 34 117 491228

19

The first PISA results provoked considerable debate in countries that have strong selection processes at the transition from primary to secondary education and a differentiated secondary education system with schools offering different kinds of curricula to what they take to be recognisably different kinds of students. These arrangements seemed in many cases to be associated with low average and widely dispersed performances as well as a strong link between students' social background and their educational achievements. These systems appeared to be systematically reproducing existing social arrangements, conferring privilege on the already privileged and denying to those non privileged.

The dramatic change in Poland's PISA results from 2000 to 2003 following the move from selective secondary schools to comprehensive ones strengthens the general conclusion about the disadvantages of early selection. (Selection becomes inevitable in the upper secondary years as students are given more choice over the subjects they might study. In any case, the PISA results are for 15-year-olds and do not speak to the impact of later selection.)

ihe reform of the School Places Allocation System in Hong Kong is not as radical as the change made in Poland but it is in the same direction and should reduce the variation between schools. It will be important to see what effect the Hong Kong reforms have on student performance levels, in particular to see whether they raise the achievement levels of lower performing students without any lowering of the achievements of high performing students and whether they increase equity.

Moving from 13 years of primary and secondary schooling to 12 years brings Hong Kong into line with almost all developed countries. It is certainly the mainstream. Just how the twelve years are split between primary and lower and upper secondary varies considerably. The important thing to monitor is how students cope with the transitions across the various structural dividing lines. Many will watch with interest in the year in which two cohorts exit from secondary education. Some will be interested to compare the experience with that of Ontario.

A major advantage of Hong Kong's change is that it can move away from the extraordinarily narrow curriculum structure of the English A Levels and adopt a broader structure like that of the various baccalaureate systems. England remains stuck with it's A Levels, sadly branded once as the gold standard, and struggles to achieve reforms that will lift retention rates.

Curriculum issues

- Focus on learning not teaching
  - Being more explicit about standards expected of students
  - Focusing on what they should know and be able to do
  - More deliberate and rigorous monitoring in many systems
- Purposes of schooling
  - Meeting demands of knowledge economy
  - Not losing sight of personal and social outcomes
- Getting the balance right
  - Expertise – which is 'domain specific'
  - Generic competencies

mcgaw group pty ltd
ABN 34 117 491228

20

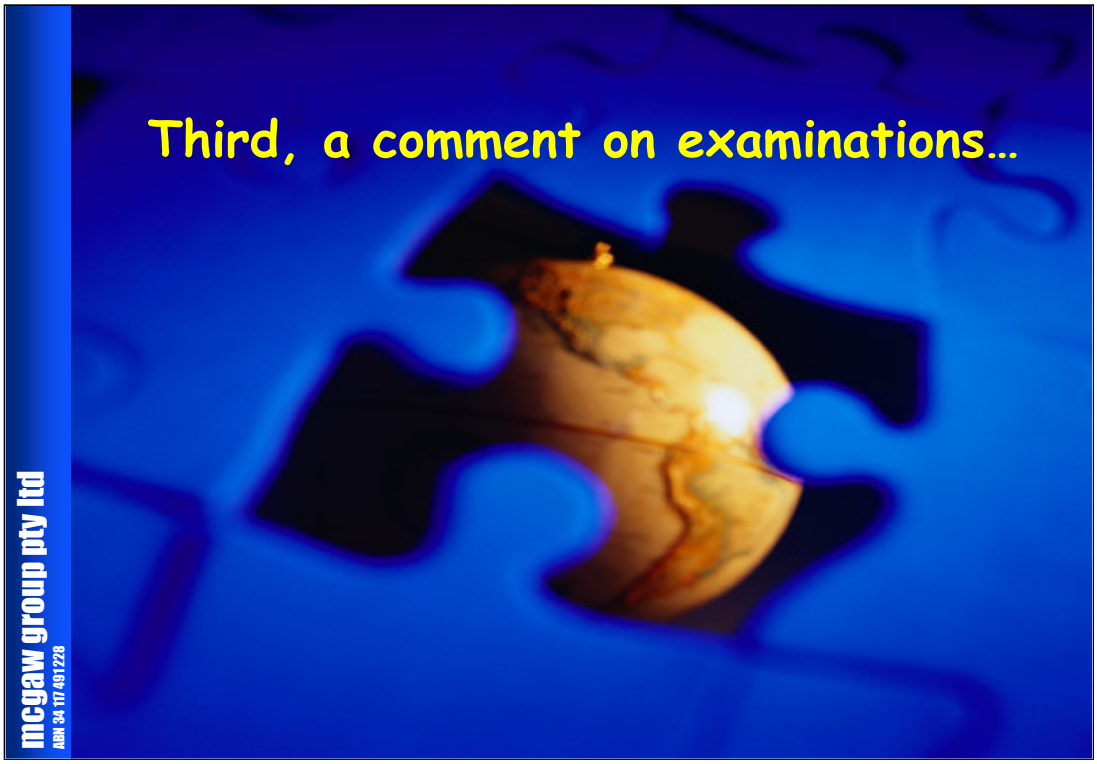
Many education systems are reforming their curricula in ways similar to Hong Kong. They are being more explicit about what they expect of students and, in the process, are generally seeking to raise expectations. The focus is on learning, not teaching, so rather than describing (or prescribing) what teachers should do, they describe what students are expected to know and be able to do as a result of their learning.

Many systems also put in place more deliberate and rigorous monitoring programs in which they monitor not only what is being achieved at the system level but also at the school level. Comparisons are made among schools to identify those that are not doing well, the comparisons being either with the group of schools enrolling similar kinds of students or with all schools using results adjusted statistically to take account of differences n the background of the students and so to estimate the value that the school is adding. Systems vary in the approach taken to poorly performing schools, ranging from the provision of special support to produce improvement to strategies for giving students the opportunity to move to other schools.

Curriculum reform must be informed by a clear view of the purposes of schooling. In some cases, concerns about the skill demands of labour markets in modern knowledge have tended to narrow the conception of purpose to a fairly utilitarian one. The result can be a preoccupation in the curriculum and in schools with basic skills of literacy and numeracy and with other 'employability skills' such as communication skills and capacity to work cooperatively with others.

It is important to maintain a balance between economic outcomes and personal and social outcomes in defining the purposes of schooling.

It is also important to get the balance right between expertise and generic competencies. Spokespeople for business and industry sometimes speak of generic competencies as though are all that matters. That is perhaps because they are talking about the competencies that they would want in all employees. If they were to be talking of what they wanted in accountants or engineers they would be more likely to say something also about specific expertise. The psychological research literature on expertise does make clear that it is domain-specific and dependent on a deep knowledge of a domain and not on a set of generic competencies that are transferable across domains. Both are needed.


mcgaw group pty ltd
ABN 34 117 491 228
Third, a comment on examinations…

.

In search of a scale for measuring individuals

- Comparing with others
  - Position in relation to the average (or norm)
  - Position in relation to distribution (e.g. percentile)
  - Improvement possible only at the expense of others
- Comparing with criteria
  - Specify learning required (Glaser, 1963)
  - Judge performance against requirements
  - Often myriad criteria so hard to make sense of results
    - Reduced to aggregations like "Satisfied 85% of criteria"
- Simultaneous scale construction and measurement
  - locate tasks on scale by difficulty
  - locate individuals on same scale by performance
  - interpret performance with reference to tasks/standards

Glaser, R. (1963), Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, **18**, 519-521.

22

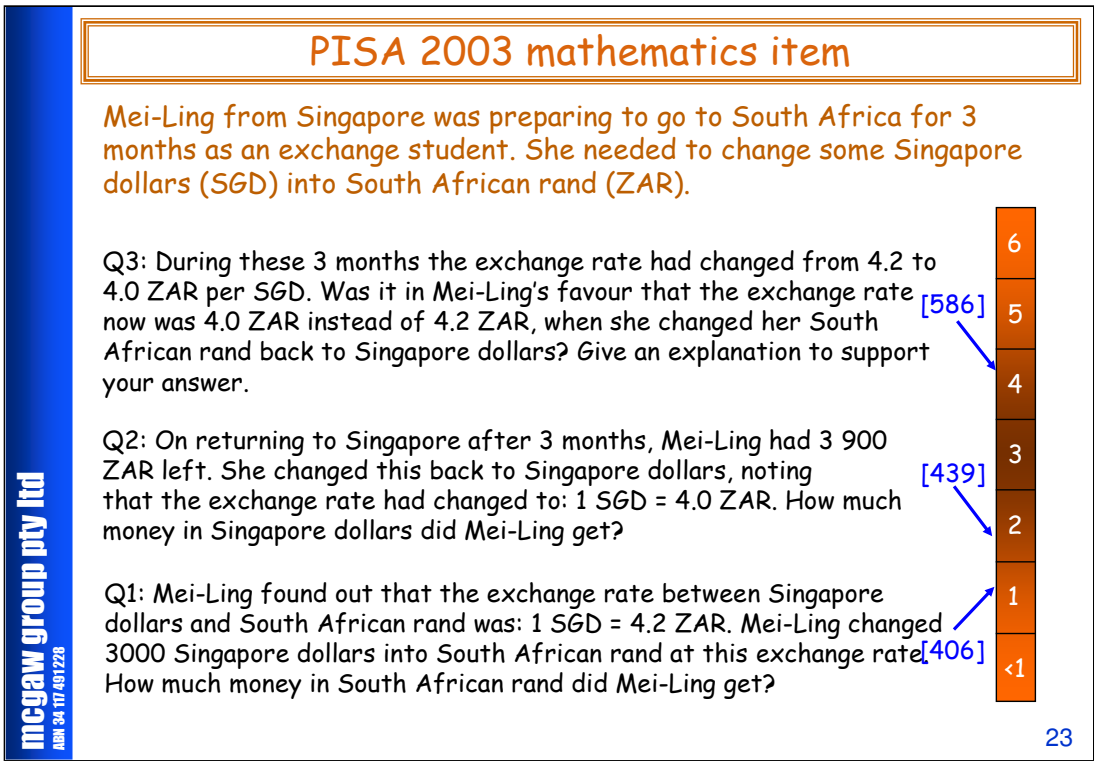
mcgaw group pty ltd
ABN 34 117 491228

In physical measurement, objective scales can be developed with which the properties of objects can then be measured. With educational (and psychological) measurement it was early accepted that, in the absence of an objective, external point of reference, the only way to interpret the performance of an individual was in comparison with the performances of others. The point of reference could be the average performance of others (i.e. the norm) of the distribution of performances of others. In the former case, an individual was described in terms of distance from the norm; in the latter, in terms of location within the distribution using measures such as percentile rank.

The most fundamental problem with this approach is that improvement can only occur in relation to others or, more specifically, at the expense of others. If all individuals were to be improving uniformly in an absolute sense, all would be interpreted as not having moved since none would have changed in relation to others. This is so fundamentally opposed to the notions of what education tries to achieve in terms of individual development that it was inevitable that a way would finally be found out of this impasse.

Robert Glaser (1963) introduced the conception of 'criterion-referenced assessment' in which individual performances would be judged against well-defined criteria. Implementations of this approach often used a myriad of quite specific criteria with no helpful way to summarise performance other than in terms of the proportion of criteria satisfied. In some cases, a satisfactory performance was taken to be to have satisfied 85% of the criteria, with no consideration of any differences between the criteria satisfied and those not satisfied.

More recent developments in psychometrics have overcome this problem and provided a means to develop well-defined scales on which tasks (or criteria) can be located in terms of their difficulty and individuals located in terms of their level of performance. This then makes it possible to interpret the performances of individuals in terms of the tasks and ultimately the scale that they define. This approach is illustrated in the following slides.



The slide above shows one task from the PISA 2003 mathematics test, in which three particular questions of different difficulty are asked. Their location of the mathematics scale, by difficulty, is illustrated.

Question 1 is located at 406 on the scale indicating that it is relatively easy and is in the range defined as Level 1 on the scale. Students performing below Level 1 would be unlikely to answer even this item correctly; those performing at higher levels would be increasingly likely to answer it correctly.

Question 3 is much more difficult as indicated by its location at 586. The scaling of the questions shows their relative difficulties. The difference between Question 3 and Question 2, in this case, is considerably greater than the difference between Questions 1 and 2. To answer Question 3 correctly a student would need to be performing in the range from high in Level 4 and above.

To fully define the mathematics scale illustrated would require further items with difficulties spread along the full range from Level 6 to Below Level 1.

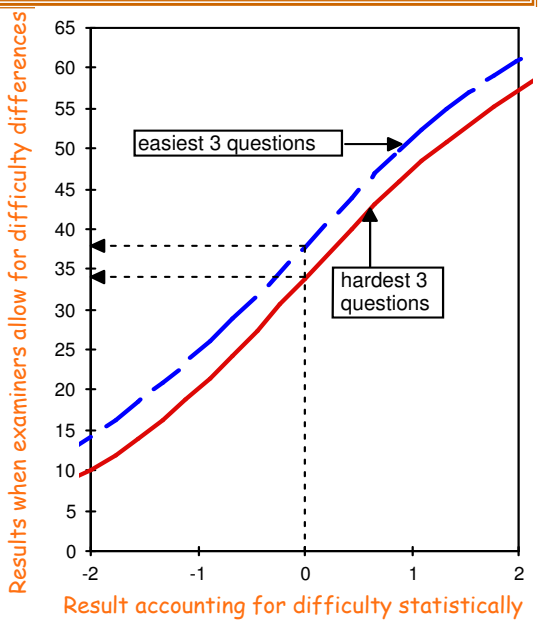
The numbers used to represent difficulty (406, 439, 586 in the cases of the three questions illustrated) are in a range arbitrarily chosen, in exactly the same way as the Fahrenheit and Celsius scales are arbitrarily chosen to represent relative differences in temperature.

Allowing for differences in question difficulty

In examinations offering choice of questions, examiners may consider question difficulty in marking. How well do they do?

The 1996 New South Wales Grade 12 Geography examination gave choice of one of 32 combinations of three questions.

Taking statistical account of question difficulty, students with the same overall performance would have received around 4 marks less from the examiners if they had chosen the most difficult 3-question combination compared with the easiest 3-question combination.



McGaw (1997), *Shaping their future: Recommendations for reform of the Higher School Certificate.* Sydney: Department of Training and Education Co-ordination, Appendix C.

24

mcgaw group pty ltd
ABN 34 117 491228

When tasks are properly calibrated on scales they can be used consistently in measuring individuals. This is not the case with standard practices in examinations where there is a choice among questions. The normal practice is to allocate marks for each response to a question and then to add up the marks. This requires the assumption that all questions are of equal difficulty or some adjustment by examiners in allocating marks to take account of differences in difficulty perceived by the examiner(s).

The appropriateness of this standard practice was evaluated by comparing its results with those from a calibration of questions by differences in difficulty in a public examination in Geography in New South Wales. In this examination, students had to answer three essay questions. There were 32 combinations of three that they might have chosen.

Students at the same overall performance level who answered the three easiest items received around 4 more marks from the standard examination marking than they would have if they had answered the three most difficult items. If the examiners had been making some adjustment in their mark allocations to take account of difficulty, this shows that they were not getting it right. Students answering the more difficult questions suffered at their hands.

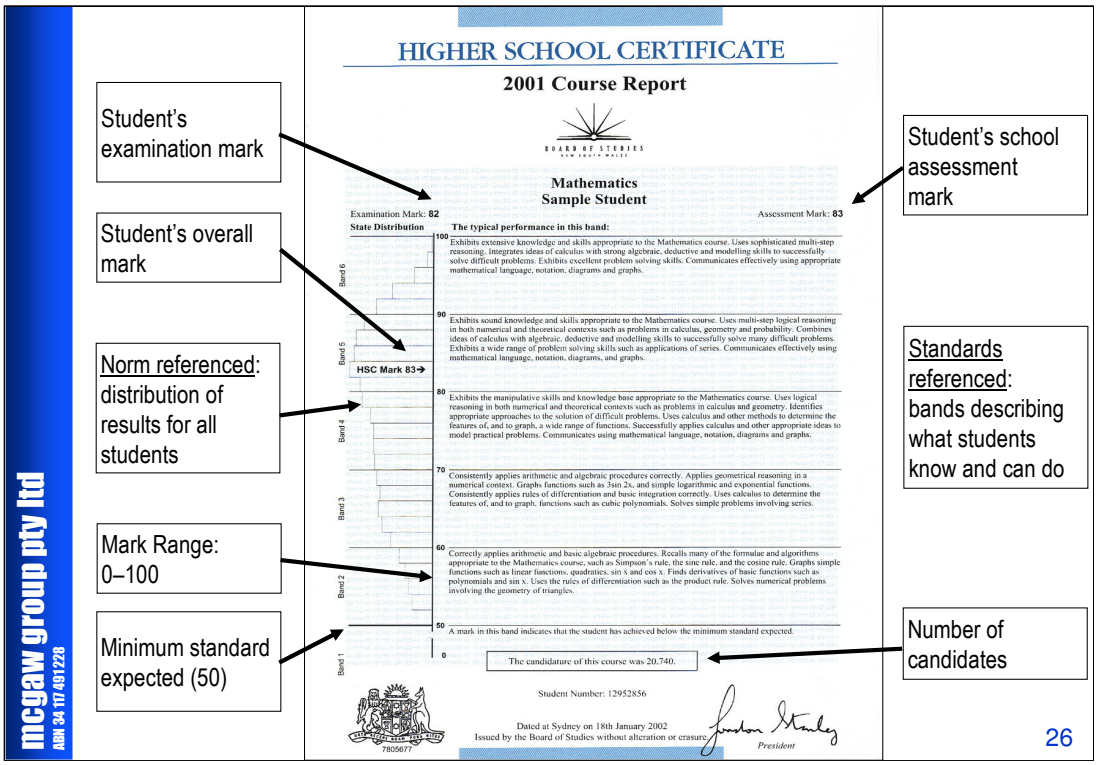
Standards and norm-referencing in examinations

- Australia (New South Wales)
  - development of band descriptors
  - 'consistent' definition of bands over years.
  - reporting with norm and criterion-referencing
- Review and recommendations for change
  - New NSW Higher School Certificate
    - McGaw, (1997). *Shaping their future: Recommendations for reform of the Higher School Certificate*. Sydney: Department of Training and Education Co-ordination
  - Scaling process
    - standards-referencing to curriculum and over-time
    - Bennett, J. (2001), Standards-setting and the NSW Higher School Certificate www.boardofstudies.nsw.edu.au/manuals/pdf_doc/bennett.pdf

mcgaw group pty ltd
ABN 34 117 491228

25

This approach to scale definition through the calibration of tasks and their use in measurement of individuals has been implemented in the end-of-secondary school examinations in New South Wales in Australia. Details can be found via the references given above.



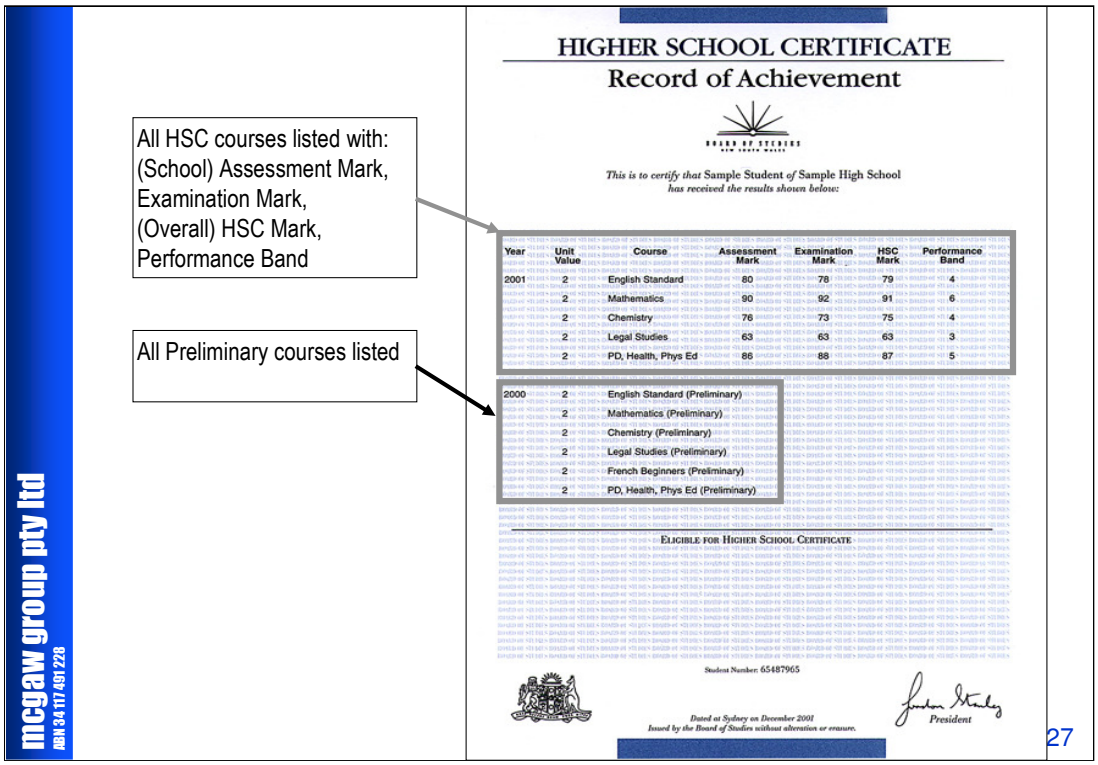
For each subject studies, students receive a course report in the format shown above. The performance scale is described at six levels and the student's overall mark locates the student on the scale. The student's competence in the subject can be described in terms of the levels on the scale. The student is likely to know and be able to do all that is described at being at lower levels on the scale than the student's performance but not likely to know or be able to do what is described at higher levels on the scale than the student's performance.

The scale has been set to range from 0 to 100 with 50 set at the point defined as the minimum acceptable standard. Performance below this on the scale would be classified as 'failing'.

This kind of interpretation of a student's performance is standards-referenced.

Use of the scale with a group of students also permits norm-referenced interpretations. On these course reports, the distribution of performances for all students is graphed on the left. This shows how the student's performances were spread along the scale and it shows where the individual student whose overall mark is shown on the scale stood in relation to other students.

The psychometric models behind this approach thus permit both standards-referenced and norm-referenced interpretations of results.



In addition to individual Course Reports, students in New South Wales also receive a summary Record of Achievement in the form shown above.

Role of school-based assessment

- Validity problem
  - External examinations limited in what they can measure
  - School-based assessment – uncertainty about who is measured
- Reliability problem
  - External examinations a limited sample of performance
  - School-based assessment can use a larger sample
  - School-based assessments may be inconsistent between schools
- On balance
  - Many systems now use school-based component
    - Use meetings of teachers to achieve consistency of marking
    - Scale school-based against external – particularly when high-stakes
  - Some use only school-based (USA, Queensland in Australia)
    - But then add some test of general skills to adjust for school differences

mcgaw group pty ltd
ABN 34 117 491228

28

Most public examination systems have been reformed to include a component of assessment undertaken in schools. This results from an, on balance, consideration of issues of validity and reliability.

External examinations suffer from invalidity because there are things that they cannot measure. Their short timeframe, for example, means that they cannot assess sustained work by students dealing with complex issues. Teachers usually see their task to be to prepare students so well for an examination that the student can provide well-rehearsed responses without doing much that is novel. School-based assessments that do not simply mirror public examination formats suffer a different form of invalidity. If the work is sustained and completed, at least in part, outside the school, then it may be difficult to discern what assistance the student has had and to what extent it is not actually the student's work.

External examinations collect only a limited sample of student work. On a different day and with a different set of questions covering the same curriculum, a student could well perform quite differently. That kind of inconsistency is unreliability in measurement. School-based assessment can collect a bigger sample of performances and so achieve more reliable measurement. If there are marked differences between schools in how marks are awarded or grades given, then that is a different form of unreliability that is avoided (or limited) in public examinations by the use of clear marking guides and supervision and monitoring of examiners by chief examiners.

The on-balance judgement made by most examining bodies now is that a mix of external and school-based assessment is appropriate but that steps have to be taken to be taken to ensure consistency of marking of school-based assessment across schools. This is sometimes done by convening meetings of teachers to achieve agreement on how samples of work would be marked. Where the examination is high-stakes, for example being used for university entrance, then consistency of school-based assessment is often sought by statistical scaling against the external examination results. That does diminish the role of the school-based assessment, of course, but the on-balance decision is that this is an appropriate price to be paid to avoid the unreliability of assessment that would be the consequence if school-based assessment awarded marks inconsistently across schools.

Some systems, such as the US and Queensland in Australia use only school-based assessment in subject areas but they add a non-curriculum based external test for 'scaling'.


mcgaw group pty ltd
ABN 34 117 491 228
Finally, a comment on post-secondary education…

.

Tertiary education

- Managing growth
  - Institutional provisions
    - Differentiated or not
      - Unified systems (so-called) – e.g. UK, Australia
      - Binary systems – e.g. Finland, Switzerland
      - Heterogeneous system – e.g. USA
    - Professional courses delayed until postgraduate stage
      - USA, Canada
      - University of Melbourne's new 'Melbourne Model'
  - Financing growth
    - Limits on public funding capacity
    - Increased requirement for private provision
      - Poor overestimate costs and underestimate benefits
      - Private contribution deferred, conditional and sometimes reduced
    - Czech Republic rejected private contribution despite
      - 5th highest completion rate of upper secondary education (of 30)
      - 29th highest completion rate of tertiary education (of 30)
      - Large salary benefit for tertiary over secondary graduates

30

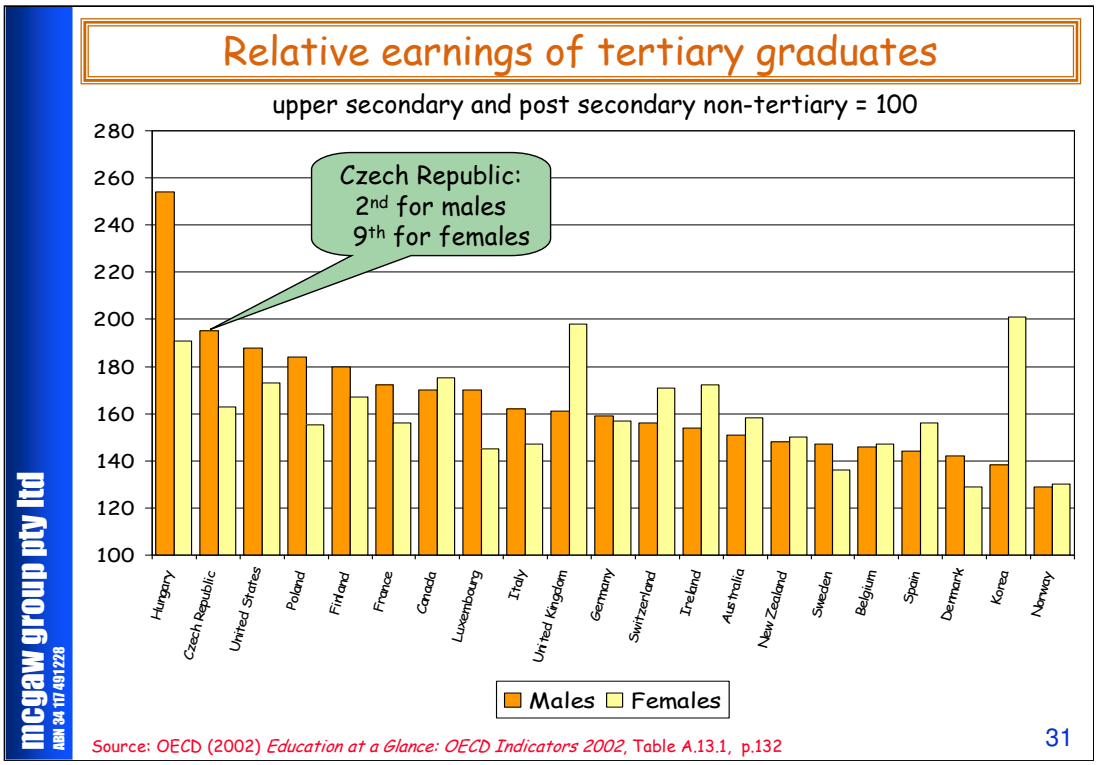
mcgaw group pty ltd
ABN 34 117 491228

With upper secondary education approaching universal provision in developed countries, post-secondary education has been expanding rapidly to the point where it is a mass system in many countries. The two biggest issues that are faced in managing this expansion are whether to differentiate the system and how to fund the expansion. Both feature highly in a current OECD study of tertiary education policy options for member countries.

The United Kingdom and Australia both had what they called binary systems with universities and technology and other institutes as distinct institutions. Both countries amalgamated these into unified national systems though various sub-groupings of institutions have emerged as a means of claiming distinctiveness again. Others, such as Finland and Switzerland, have maintained a strong distinction between universities and institutes of technology. The USA could be said to have no 'system' since there are many different kinds of post-secondary institutions with different missions.

A distinguishing feature of the US, and to some extent the Canadian systems, are that professional preparation in fields such as medicine and law are delayed to graduate schools and are build upon a completed undergraduate course. The University of Melbourne is currently moving to adopt this model in an extensive way. The Universities of Sydney and Queensland and Flinders University have already done it for medicine. Delaying entry to high-demand professional courses to graduate-level reduces the pressure on students in the final year of secondary education and on the selection role of end-of-secondary education examinations.

With other competing priorities and pressure to contain public expenditure, many countries look to private contributions to fund some of the growth. There are equity concerns if access to tertiary education becomes dependent on the personal or family wealth of potential students so various strategies have been developed to minimise the barriers faced. The Australian Higher Education Contribution Scheme requires students to make a financial contribution but only after they have graduated and only if their income is above a threshold. The government thus carries the risk of there being insufficient financial benefit for the individual from tertiary education. That helps to keep the barrier low for poor students who tend to overestimate the cost and underestimate the benefits. The Australian scheme has been adopted in New Zealand, England and Scotland. It was considered and rejected in the Czech Republic.



The Czech Republic has the 5th highest completion rate for upper secondary education among the 30 OECD countries (as shown in Slide 18) but only the 29th highest completion rate for tertiary education. The reason for the discrepancy is that tertiary education is a restricted option available to only an elite stream among those completing upper secondary education.

The argument against the adoption of the model for a contribution from students towards the costs of their tertiary education was that tertiary education is a public not a private good, with the entire society benefiting from the tertiary education from those who obtain it. While there is no doubt that there is a public benefit it is clear that there are also substantial private benefits as the figure above shows.

The figure shows the pre-tax salary differences between tertiary education graduates and those who have completed only secondary or non-tertiary, post-secondary education, the average annual earnings for the latter group indexed to 100 in the figure, separately for males and females.

In all countries, tertiary graduates earn more than those whose formal education ends at upper secondary level or post-secondary, non-tertiary level. The premiums in the Czech Republic are 95% for males (2nd highest among the 21 OECD countries for which data are available) and 63% for females (9th highest). In some countries the premium for males is lower than that for females. This does not mean women earn more than men in those countries, of course, only that the benefit for tertiary graduates, compared with non-graduates, is greater for women.

It is quite clear that there is a private benefit from the completion of tertiary education. Fuller analyses of the lifetime benefits, taking account of costs of completing tertiary education and the income tax and post-tax income benefits are provided in OECD's annual *Education at a Glance* for those countries for which the necessary data are available. These analyses confirm the picture provided by the simpler analysis in the figure above.

Role of qualifications frameworks/systems

- Continuing learning has economic and social benefits
  - Rising skill demands in many economies
  - Risk of social polarisation if those ahead move further ahead
- Contribution of qualifications systems
  - Can they motivate learning by promise of recognition
    [See OECD (forthcoming) *Qualifications systems: Bridges to lifelong learning*]
  - Countries need, among other things, to consider:
    - Improving needs analysis to ensure that qualifications are up to date
    - Recognising non-formal and informal learning
    - Providing credit transfer
    - Clarifying learning pathways
    - Improving information and guidance about qualifications systems
    - Improving use of qualification in recruitment
    - Ensuring qualifications are portable
    - Communicating benefits of learning for qualification

mcgaw group pty ltd
ABN 34 117 491228

32

Beyond initial education, there is a growing requirement for individuals and countries to see further education as an integral part of lifelong learning. Changing demands in the labour market as economies alter is one driving force. The importance of minimising social polarisation as some are left behind with inadequate skills for sharing in the benefits of a knowledge economy is another.

Many OECD countries believe that a well constructed and managed qualifications framework could help to motivate individuals to continue learning. A full study of the nature and role of national qualifications systems in OECD countries is about to be released by the OECD as *Qualifications systems: Bridges to lifelong learning.*

Among this report's recommendations on actions that countries might take to improve the impact of their national qualifications systems are the eight listed above.

In summary

- Hong Kong's education system is currently
  - High quality (at least at mid-secondary)
  - Relatively high equity
  - But structured in ways that reflect existing social structures
- Hong Kong's education reforms
  - Are impressive in their intent and scope
  - Are well targeted to achieve their aims
  - Are in the mainstream
  - But need to monitored carefully to ensure they do
  - And should be reported internationally so others can learn from your efforts and your successes.

mcgaw group pty ltd
ABN 34 117 491228

33

Hong Kong has a high-quality and relatively high-equity education system, at least up to the mid-secondary years for which international, comparative data are available to provide the supporting evidence. The organisation of school enrolments, however, does structure schooling in ways that, to some extent, reflect existing social arrangements. There are systems where this relationship is much stronger and of which it could be said that they play an essentially social reproductive role, reproducing existing arrangements by conferring privilege where it already exists and denying it where it does not. That is not true of Hong Kong and the changes in Hong Kong's School Places Allocation System are likely to make it less true. That is a matter that should be investigated as the change works through the system.

Hong Kong's current education reforms are impressive in its intent and scope and they are also well targeted to achieve their aims. The changes are certainly in the mainstream of international developments.

It is important that the implementation of the reforms continue to be monitored carefully to ensure that their objectives are realised. It is also important that they be reported internationally so that others can learn from your efforts and your successes.

Thank-you

Contact

barry.mcgaw@mcgawgroup.org
bmcgaw@unimelb.edu.au

OECD education website

www.oecd.org/edu

mcgaw group pty ltd
ABN 34 117 491 228

34