

# Strengths and weaknesses of Queensland's OP system today

Reg Allen

October 2013

Contents:

1	Purpose .....	4
2	Introduction .....	4
3	Key elements of the design of the current OP system .....	6
3.1	Key element : a measure of overall achievement in senior secondary studies.....	6
3.2	Key element: the main purpose of the OP is its use for tertiary entrance .....	6
3.3	Key element: students can choose from a wide range of subjects .....	7
3.4	Key element: to get an OP students take five or six subjects – breadth without specification of essential subjects .....	8
3.5	Key element: the school is the decision maker .....	8
3.6	Key element: Order and gaps (SAIs) .....	9
3.7	Key element: two stage scaling model .....	9
3.8	QCS test as a reference measure .....	9
3.9	FPs .....	10
3.10	Refinements to deal with anomalous situations .....	10
4	Key assumptions .....	11
4.1	Assumption: there is a well-defined cohort completing year 12 at a given time.....	11
4.2	Assumption: there is a reasonable range of participation within school .....	11
4.3	Assumption: decisions about order and gaps in SAIs are based on identifiable features of student work.....	12
4.4	Assumption: SAIs are not intentionally correlated across subjects within school .....	12
4.5	Assumption: SAIs and Levels of Achievement are related but different .....	12
4.6	Assumption: achievements in different subjects have enough in common with each other to define a single construct .....	13
4.7	Assumption: the components of each school data set have enough in common with each other to define a single construct that relates sufficiently well to the state-wide construct that underpins the OP 13	
4.8	Assumption: QCS relates well enough with the construct – within schools and overall.....	13
4.9	Assumption: the system can be explained and understood.....	14
5	Strengths of the current OP system:.....	14
5.1	The key input – SAIs – uses what teachers can do well in identifying fine-grained distinctions between achievements.....	14
5.2	The school-decision model puts the important decisions where they should be .....	15

5.3	Triangulation: SAIs, LoAs and the QCS test.....	16
5.4	Allows wide range of subject combinations from a wide range of subjects .....	16
5.5	Sophisticated processes.....	17
6	Weaknesses of the current OP system: .....	17
6.1	In principle weaknesses: .....	17
6.1.1	The OP is focused on university entrance.....	17
6.1.2	Breakdown of fundamental assumptions.....	17
6.2	In practice weaknesses .....	19
6.2.1	The competitive academic curriculum.....	19
6.2.2	Changing patterns of participation .....	19
6.2.3	Decline in participation – alternative means .....	34
6.2.4	Bonus schemes and FPs .....	34
6.2.5	Pressures to closely align Levels of Achievement and SAIs .....	34
6.2.6	Public discourse .....	34
6.2.7	The need for change .....	35
7	Conclusions and implications.....	35

# 1 PURPOSE

---

The purpose of this paper is to inform the Queensland Review of Senior Assessment and Tertiary Entrance Processes by describing the strengths and weakness of the 'OP' system in Queensland as it is in the opening years of the second decade of the twenty-first century.

# 2 INTRODUCTION

---

The paper is intended to provide a basis for consideration of possible changes to the current system. It therefore focuses on those underlying features that set the fundamental shape of the system.

Any significant change, rather than incremental improvements and refinements, will involve change to one of more of these underlying features. Such changes are likely to affect other aspects of the Queensland system. Equally, if the review of other aspects of the Queensland system changes, directly or indirectly, any of these underlying features, the OP system will change, whether that change is wanted or not.

A chapter of the Pitman Report (2001, pp 106 – 117), written by the present author, identified the need to plan for changes in the then current system (since key assumptions on which it was based were likely to be less well met in practice) and key requirements for a system aligned with the more diverse ways of completing senior secondary studies. The published version of this is chapter summarised the situation as follows:

*The assessments from which OPs are determined, SAIs, are based on the notion of a two-year cohort of senior students at a school completing study in a subject at the same time. The greater the mismatch between this assumption and practices the less workable the system. This is a fundamental rather than a technical difficulty. That is, as the trend towards more diverse approaches to senior studies strengthens (and it will be reinforced by the New Deal proposals, precisely because this diversity of approaches is needed to increase effective participation), the present OP system will have to change in its fundamentals; the within- school, within-subject comparisons that form its input data will become increasingly unviable. OP calculations are done each year for the group of students completing Year 12 that year, finishing two-year courses of study in subjects at the same time. The present system works because it is reasonable to ask a teacher to provide a class rank order at the end of the course and because there is a state-wide cohort of Year 12 students completing their courses at the same time. It will not continue to work in its present form when there is no such class, no group of students completing the course at the same time, and no Year 12 cohort across the State all completing their Year 12 courses at the same time.*

*This will not happen overnight — it is the result of a longer term trend. There is opportunity to develop the new system that is required. This paper sets out some options for developing a new system consistent with the New Deal proposals and for managing the transition period. (Allen 2001 in Pitman 2001, page 110 – 111)*

The development of a different system has not yet occurred. The need for it has increased, not diminished, to the extent that it is important to design and plan for significant changes. A later paper in this series will identify feasible options. Transition will also require careful design and management.

The current system has, as described below, both strengths and weaknesses. The latter mean that change is important and necessary. On the other hand, the strengths of the current system mean that the need for change is not so urgent that students and parents should be concerned about the fundamental fairness of OPs in 2013 and 2014.

The OP system was developed, from its initial design in 1987 through to more-or-less its current form in the mid-1990s, to deal with the competing considerations of

- pressures arising from competition for places in university courses
- basing university entrance on results in senior secondary studies
- reporting achievement with a precision based on the properties of the data.<sup>1</sup>

Competition for university places occurs when there are more qualified and capable applicants than there are places – an excess of demand over supply. Competition is lessened when there is an increase in supply and or a decrease in demand. In recent years, it appears that such changes have been happening in relation to many courses and universities. There is still serious competition for some courses at some universities, however. And the resolution of this competition has ‘backwash’ effects on what students do in senior secondary studies. It would be a mistake to imagine that it is not necessary to design a tertiary entrance system that takes this into account and meets the needs of students and universities for fair, transparent and efficient selection processes.

The current OP system is part of the Queensland system of externally-moderated school-based assessment. A recent paper in a continuing World Bank series providing data and evidence on what matters most in driving quality in education described and analysed the factors that created and maintained the Queensland system. In this paper<sup>2</sup>, the author observes that

*The Queensland approach requires consensus among key stakeholders on the following beliefs or guiding principles: that teachers are best qualified to judge the achievement of their students; that assessment activities should never be separated from curriculum and instruction; and that the construct and consequential validity of assessment results (and the impact on learning) should take priority over a narrow focus on psychometric concerns about reliability and equating and the value of standardized testing. (Allen 2012, p. xiii)*

The current OP system reflects these guiding principles. A challenge for the design of a changes to the OP system is whether to retain these principles, while changing its design and practice, or to discard one or more of them.

---

<sup>1</sup> This is why the OP is in broad bands. It is possible to provide scores with many digits – 98.456343, for example. But the data do not support such a degree of apparent precision or, more importantly, the fairness of decisions based tiny differences.

<sup>2</sup> Allen, Reg. 2012. *Australia - Developing the enabling context for school-based assessment in Queensland, Australia*. Systems Approach for Better Education Results (SABER) student assessment working paper ; no. 6. Washington D.C. - The Worldbank. <http://documents.worldbank.org/curated/en/2012/12/17191778/australia-developing-enabling-context-school-based-assessment-queensland-australia>

We focus first, therefore, on the key elements of the current system, before turning to its fundamental assumptions, its strengths and its weaknesses.

### 3 KEY ELEMENTS OF THE DESIGN OF THE CURRENT OP SYSTEM

---

Accounts, in summary and in detail, of the OP system designed for audiences of teachers, students, parents and the community are readily available on the Queensland Studies Authority web-site. An overview of the Queensland system, including the OP, written for an international audience not familiar with local terms and ideas, appears in Allen (2012).

This section focuses on the key elements that underpin the design of the current system. In this context, understanding the current system as a system requires understanding the conceptual basis of each of these elements. For this purpose, the concept is what matters, the operational details are not important.

#### 3.1 KEY ELEMENT : A MEASURE OF OVERALL ACHIEVEMENT IN SENIOR SECONDARY STUDIES

The OP (Overall Position) is a rank order from 1 (the highest) to 25 based on students' overall academic achievement in senior secondary studies<sup>3</sup>. It is not a 'score' in the sense of 'number out of', despite the frequent references in the media to 'OP score'. It was designed to be approximately comparable from year to year – so that a student with an OP 2 from one year can be considered to have achieved – overall – more highly than a student with an OP 4 from another year.

Overall academic achievement is not subject-specific achievement but is related to it – just as achievement in a pentathlon is overall achievement in that contest, related to but not the same as achievement in each event. This concept seems to be readily grasped in the context of sporting contests but is often misunderstood in discussions about the scaling processes used in Australia to derive measures of overall achievement from subject results.

As acknowledged in Viviani(1990), having an OP together with results in individual subjects is a policy decision: designing the system to avoid putting too much pressure on any one component. As Allen (2012) notes

*Too much pressure happens when the results from a single component – an examination, a moderation system – are relied on alone for high-stakes decisions. The consequences of such excess pressure can show up in different ways: for external tests it can lead to a preference for reliability over validity; for moderation systems it can lead to malpractice of one kind or another.*  
(Allen 2012, p 8)

#### 3.2 KEY ELEMENT: THE MAIN PURPOSE OF THE OP IS ITS USE FOR TERTIARY ENTRANCE

The OP, which is a measure of overall achievement at one stage of education, is described as being essentially for the purposes of selecting students for tertiary education – a process sometimes called

---

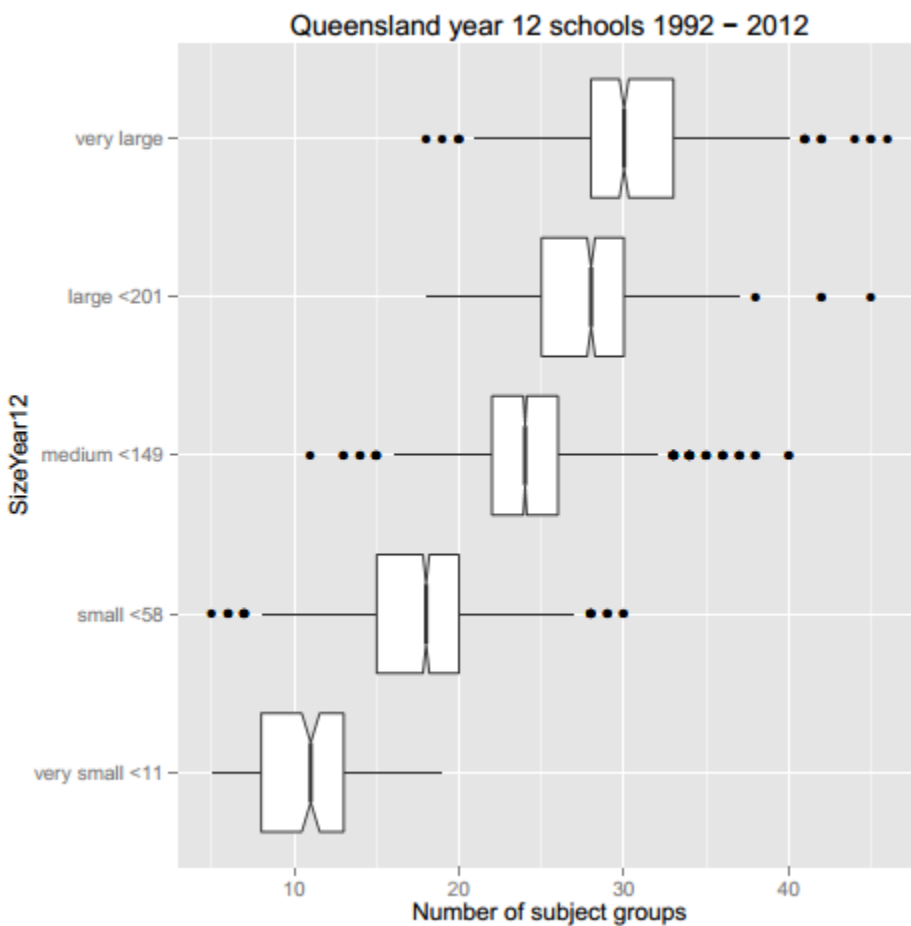
<sup>3</sup> The term 'overall academic achievement' is used here in the sense that a combination (an aggregate or an average) of results across a student's different subjects represents a measure of achievement overall. Grade-point average is an example of a measure of overall achievement, one based on the assumption that grades are comparable across subjects without any scaling. It is well known that grades do not have this property.

university entrance. This may seem so obvious that it is not worth remarking on. However, treating tertiary entrance as the central purpose of the OP regardless of the other uses to which it is put affects the sorts of subjects included and excluded and the rules for eligibility.

### 3.3 KEY ELEMENT: STUDENTS CAN CHOOSE FROM A WIDE RANGE OF SUBJECTS

Students can choose from a wide range of subjects – there are around 60 different subjects with results that can be counted towards an OP. No individual school offers this many – a total of about 40 is the upper limit for schools with a large or very large year 12 cohort<sup>4</sup>.

As can be seen from the following chart<sup>5</sup>, there is a clear tendency for larger schools to have more subjects<sup>6</sup>. However, there are some very large schools with fewer subjects than some schools with small year 12 cohorts. There are some very small schools with a lot of subjects.



<sup>4</sup> This is the count of year 12 students, not the count of OP-eligible students. Data sets provided by the Queensland Studies Authority.

<sup>5</sup> Analyses and charts in this report done with R: R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

<sup>6</sup> If the notches in two boxplots do not overlap it is likely that there is a significant difference in the medians.

Further analysis of the data suggests that there is a tendency (once the size of the year 12 cohort has been taken into account) for the number of subject groups per school to increase

- over time – there is a marked difference between the 1990s and today
- with region – the more urban regions
- with the QCS mean and spread of the school group
- with an estimate of current average socio-economic status.

### **3.4 KEY ELEMENT: TO GET AN OP STUDENTS TAKE FIVE OR SIX SUBJECTS – BREADTH WITHOUT SPECIFICATION OF ESSENTIAL SUBJECTS**

While there are many possible subjects, an individual student takes between five and six subjects (occasionally more). In the last ten years, the proportion taking five only has fluctuated around 24 per cent, dipping below this in 2009 to 2011.

Calculating a measure of overall achievement common across different students requires that each student's set of specific subject results covers a range of subjects. One way to do this is to define some explicit requirements for diversity and balance in each student's choice, perhaps by specifying some compulsory subjects or by specifying lists of subjects and requiring each student to choose one from each list.

It is however a feature of the current OP system that requiring each student to have at least five results in different subjects, together with long-standing assumptions by schools, teachers, parents and students, have meant that no rules specifying restrictions on combinations of subjects have been needed – most students have taken a reasonably broad range of subjects. While there has been a decline in the last decade in the proportion of students with four semesters of English, this has been from 99.5% to 98.9%. The proportion with four semesters of mathematics has increased over the same period from 96.7% to 98.3%, a trend that may be associated with changing patterns of participation.

Generally, this means that most combinations of subjects taken by an individual student are not so narrow as to be repetitions of essentially the same subject – something that is necessary for the construction of a measure of overall achievement that is reasonably common across students<sup>7</sup>. The introduction of 'extension' subjects has the potential to change this. The numbers of students with results in one or more units of credit<sup>8</sup> in extension subjects has more than doubled over the last decade, from less than three per cent in 2003 to 6.5 per cent in 2012.

### **3.5 KEY ELEMENT: THE SCHOOL IS THE DECISION MAKER**

The current OP system sits within the Queensland system of externally moderated school-based assessment.

---

<sup>7</sup> A measure of overall achievement can be thought of in terms of the first principal component from the covariance matrix of all of the subject achievement indicators. If this covariance matrix includes multiple entries for essentially the same subject, the first principal component will align with this subject. In practice, of course, the fact that subjects are not selected at random means that estimation of this covariance matrix is difficult. The subject selection mechanism is not random, but may not be explained in terms of a single, simple process.

<sup>8</sup> Nearly all (97.5%) students with results in extension subjects have two units of credit in these subjects.



Outside Queensland, the idea of school-based assessment is often confused with *classroom assessment*. Classroom assessment usually connotes relatively informal assessment at the individual class level by an individual teacher. Where this form of assessment is used for summative purposes it is well known to lead to ‘grade-inflation’ and to actual or perceived exercise of caprice and prejudice. School-based assessment in the Queensland assessment system connotes a program in each subject of systematic and planned assessment activities for which the *school is* responsible and that are used for state-wide certification. The school, not its individual teachers, is the accountable agency. It signs off on the assessment program and is responsible for ensuring its implementation.

Within this environment, the input data for the determination of OPs is determined by the school.

There are pressures to remove the human element from high-stakes decisions in many fields, including education. In the OP system, these manifest as attempts to find ways to make school decision-making about the input data more mechanical rather than judgmental.

### 3.6 KEY ELEMENT: ORDER AND GAPS (SAIs)

In the OP system, consistent with the principle of the school as the decision-maker, the input data takes the form of within-school within-subject interval scales<sup>9</sup>. These are known as Subject Achievement Indicators (SAIs). SAIs provide a finer-grained set of comparisons (a 200 point scale) than the five point scale provided by the standards-based subject results (levels of achievement – LoAs). SAIs are comparisons of *relative* achievement – enacting an assumption that teachers can make fine-grained comparisons of the work of students they have taught. The standards-based LoAs are, of necessity, a coarse scale, requiring teachers to make comparisons of student work with a set of abstractions –the definitions of standards.

In the 1980s, there were attempts to use paired-comparisons approaches to determining SAIs. The technology may now be emerging to support this sort of approach.

### 3.7 KEY ELEMENT: TWO STAGE SCALING MODEL

SAIs make comparisons within a subject within an individual school. Putting SAIs from different subjects together requires that these be placed on a common scale (overall achievement).

In the current OP system this is done in two stages. First, there is a within-school scaling that puts all the SAIs in the school on a common scale so that they can be added up to give a single composite – overall achievement within that school.

Secondly, there is a between-school scaling that puts the individual school composites on a common scale – overall achievement across the state.

### 3.8 QCS TEST AS A REFERENCE MEASURE

The Queensland Core Skills Test (QCS Test) is a test of general academic achievement taken by all (eligible) Year 12 students in late August or early September.

---

<sup>9</sup> Interval scales have order and gap properties but no zero.

The principal purpose of the test is to gather group (school and subject class) information (measures of central tendency and spread) to put SAIs onto a common scale. Students' rankings or scores in the QCS test are not combined with their subject results.

The QCS Test is not an external examination in the sense that it is used for individual high stakes certification: students receive individual results but, unlike external examinations, these are not used for any high-stakes decisions for the individual – two individuals with identical subject results and the same tertiary entrance rank can have very different QCS results.

As noted above, the underlying construct of the OP is 'overall academic achievement', not subject-specific achievement. The underlying construct of the QCS test is, correspondingly, overall academic achievement. The test construct is given in more detail in terms of a set of 49 common curriculum elements.

When QCS test group results are used to compare the academic strength of, for example, a school's group of students doing Physics with its group doing French the underlying construct is neither French nor Physics but general or overall academic achievement as estimated from a test of common curriculum elements.

### 3.9 FPs

Field Positions (FPs) show a student's rank order position (on a one to 10 scale, with one being the highest) based on a unequally weighted combination of scaled SAIs<sup>10</sup>. The weights for subjects reflect judgments about the emphasis in each subject on particular types of knowledge and skills.

FPs were designed to provide additional information where needed to distinguish students with the same OP, not as the only means for selection. FPs are therefore only intended to be used at the margin and hence only a few selection decisions will involve the use of FPs. The lower the selection pressures, the smaller the number of decisions likely to be made with FPs.

### 3.10 REFINEMENTS TO DEAL WITH ANOMALOUS SITUATIONS

Over the years since Queensland first derived measures of overall achievement (then called a TE Score) from school-based decisions a wide range of refinements have been added to deal with anomalous situations. There are procedures to deal with the challenges posed by

- subject groups that are small (ten or fewer) – the scaling model makes no sense when there is only one student doing a subject in a school
- having different procedures for these "*small subject groups*"
- the impact on the validity of scaling procedures of students with anomalous QCS results
- potentially inappropriate practices in determining SAIs
- cases where a student's OP is very much lower than the OP of other students with the same levels of achievement in the same subjects.

---

<sup>10</sup> In the determination of OPs, all subjects are given the same weight. Public perception sometimes confuses the design and operation of scaling with a weighting process whereby some subjects 'count' more than others.

## 4 KEY ASSUMPTIONS

---

The purpose of describing the following key assumptions is to establish a framework within which the major strengths and weaknesses of the current system can be understood and which will help to understand the likely impact of any proposed changes to the OP system.

### 4.1 ASSUMPTION: THERE IS A WELL-DEFINED COHORT COMPLETING YEAR 12 AT A GIVEN TIME

The current system is built around an assumption that students complete two years of full-time study culminating in the completion of year 12, a tertiary entrance result and transition to further study. The idea of a state-wide rank order – a feature of tertiary entrance ranks in Australia – fundamentally rests on a snapshot at a point in time, where the overall achievements of a set of students are placed in order. While there are procedures in Australia (including Queensland) designed to make these rank orders comparable from year to year, the starting point is the comparisons made in a particular year. These start with an individual student's subject achievements, put these onto a within-year comparative scale, put these comparative scales onto a between-subject scale, combine these 'scaled' values into an aggregate and use this aggregate as the basis of a rank order.

Pressures on this assumption include

- public expectation that there is or ought to be some standards-based process for comparing academic merit
- universities' moving away from the idea of the academic year
- students' expecting to move in and out of post-year 10 study at different times and to be able to study part-time.

In the last decade, the proportion of OP eligible students completing studies over three years has remained at a very low level of around one per cent.

A further aspect of this assumption is the idea of a 'year 12' cohort at a school – the OP system assumes that at each school there is a body of students completing year 12 through study at that school. Widespread use of on-line courses, where the teaching and assessing is outside the school and of external examinations puts pressure on this assumption.

### 4.2 ASSUMPTION: THERE IS A REASONABLE RANGE OF PARTICIPATION WITHIN SCHOOL

The calculations that underpin the OP system are designed to estimate overall achievement from sets of relative subject achievements. The scaling of the input data, which is in the form of interval scales, uses estimates of the location and the spread. If there is no variation in the input data (every student has the same result), there is no spread and the processes break down completely. The closer to this situation in practice, as the group of OP-eligible students becomes more select, the greater the uncertainty introduced into the validity of the scaling processes. For example, the use of QCS results to compare two groups of students in terms of overall achievement becomes less valid the less variation there is in the achievement of one or both of these groups – the observed variance and co-variance becomes less and

less good estimates of the 'true' variance and co-variance. This is not of merely theoretical importance – a key principle of the OP calculations is the idea that an individual's OP should depend on that individual's achievement and not on membership of a group.

#### 4.3 ASSUMPTION: DECISIONS ABOUT ORDER AND GAPS IN SAIs ARE BASED ON IDENTIFIABLE FEATURES OF STUDENT WORK

The OP system assumes that decisions about SAIs (the input interval scales) are based on differences in the work students have done and on no other considerations of any kind, whether these are judgments of a student's 'real' ability that has not shown up in the work, some notion of what an individual student needs or some extrinsic factor.

#### 4.4 ASSUMPTION: SAIs ARE NOT INTENTIONALLY CORRELATED ACROSS SUBJECTS WITHIN SCHOOL

Since SAIs are to reflect relative achievement in subjects, there should be no relationship between SAIs across subjects other than that attributable to students' individual academic ability and commitment to study.

More formally, the SAI of the  $i^{\text{th}}$  student in the  $j^{\text{th}}$  subject could be modelled in terms of an overall ability factor modified by a factor reflecting particular strengths/weaknesses in and commitment to achievement in that subject plus an error term:

$$SAI_{ij} = overall_i + subject_{ij} + \varepsilon_{ij}$$

The OP system assumes that these  $\varepsilon_{ij}$  are uncorrelated across subjects and students.

Suggestions that a school is 'manipulating' its SAIs to successfully advantage some of its students are in effect suggestions that there are processes within the school that involve a breakdown of this assumption. Note that, given the two stage scaling model, any such process is essentially a zero sum one – although it might involve compensating for the advantage of a few by spreading the loss among many. Whether such manipulations occur, whether they have any significant impact or whether they are merely unfounded rumour is less important. Monitoring processes were introduced during the 1990s to identify any instances of the breakdown of this assumption and to take corrective action where required. Such processes, however, come with costs and consequences for other aspects of the system.

#### 4.5 ASSUMPTION: SAIs AND LEVELS OF ACHIEVEMENT ARE RELATED BUT DIFFERENT

In design, a Level of Achievement (LoA) is a broad brush result expressing a student's achievement in terms of state-wide standards. An SAI compares a student's achievement with that of other students doing the same subject at the same school. It is assumed, therefore, that there is a relationship between LoAs and SAIs within a school – students with a higher LoA will have higher SAIs. In practice, the system assumes a somewhat closer relationship than this. Just how close presents a challenge to this assumption: requiring SAIs to be modelled from LoAs, to represent each LoA with the same interval and to be evenly spaced within each LoA creates a tension between the within-school within-subject purpose of SAIs and the state-wide comparability of LoAs. A mechanical relationship between SAIs and LoAs reduces the role of the school as decision-maker and places too great a reliance on the precision of state-wide moderation.

#### 4.6 ASSUMPTION: ACHIEVEMENTS IN DIFFERENT SUBJECTS HAVE ENOUGH IN COMMON WITH EACH OTHER TO DEFINE A SINGLE CONSTRUCT

A key requirement of any process determining an estimate of overall achievement in senior secondary studies is that the input data (results in subjects) have enough in common with each other to define a single construct. This condition is met with moderate covariance between subjects. A subject in which achievement has little, no or a negative relationship with achievement in other subjects should in principle not be included in the OP calculations. It is important for practical purposes that there are not very many of these.

#### 4.7 ASSUMPTION: THE COMPONENTS OF EACH SCHOOL DATA SET HAVE ENOUGH IN COMMON WITH EACH OTHER TO DEFINE A SINGLE CONSTRUCT THAT RELATES SUFFICIENTLY WELL TO THE STATE-WIDE CONSTRUCT THAT UNDERPINS THE OP

The first of the two stages in the scaling model for OPs – within-school between-subject then between-school – rests on an assumption that the general condition for creation of an estimate of overall achievement applies sufficiently well in each school and that this construct can be aligned in the second stage of scaling with that in other schools.

The sparse nature of the data in most schools<sup>11</sup> means that it is not easy to do much more than very basic checks on the validity of this assumption. However, if there is not a reasonable range of participation in the school (see assumption 4.3) some of the necessary conditions *a priori* for this assumption will not be met.

#### 4.8 ASSUMPTION: QCS RELATES WELL ENOUGH WITH THE CONSTRUCT – WITHIN SCHOOLS AND OVERALL

The way that the QCS Test is used in the scaling process means that the relationship between QCS scores and overall achievement has to be adequate at only the group, not the individual, level. It needs to be adequate at both a within-school level – since this is the level at which results in different subjects are put onto a common scale – and a between-school level, since this is the level at which overall achievement in different schools is put onto a common scale.

The system has an independent measure for overall achievement within each school<sup>12</sup>. The relationship of these two measures provides an approximate indirect indicator of the relationship between QCS scores and the underlying overall achievement construct.

The relationship of QCS scores and average level of achievement provides an indicator of the adequacy of QCS scores as an estimator of overall achievement – and hence, by inference, its adequacy for its role

---

<sup>11</sup> in the average school about 100 year 12 students have results in six subjects from twenty-three subject groups, so any estimates of the subject covariances must take into account the extensive missing data, data that are not missing at random.

<sup>12</sup> This uses the SAIs as providing a set of paired comparisons. A simple indicator of relative achievement is then derived from these comparisons, based on a method devised by HA David.

in the second stage of scaling, putting within-school results onto a common state-wide scale. This assumes, of course, that exceptional cases are just that – exceptional.

#### 4.9 ASSUMPTION: THE SYSTEM CAN BE EXPLAINED AND UNDERSTOOD

The current OP system is a mature system – it has many details developed over time in response to the importance of deriving estimators of overall achievement in ways that are reasonably fair to the individual and align with policies for matters such as appropriate backwash effects on the senior secondary curriculum, diversity and flexibility in students’ choice of subjects and locating key decisions about students’ achievements with those best placed to know and understand these achievements in a full and rounded way.

All this, however, presents a challenge in building community understanding of and confidence in the system. Paradoxically, there can be more community confidence in a system whose technical details are not defensible but is thought to be simple and straightforward – and therefore fair.

## 5 STRENGTHS OF THE CURRENT OP SYSTEM:

---

### 5.1 THE KEY INPUT – SAIS – USES WHAT TEACHERS CAN DO WELL IN IDENTIFYING FINE-GRAINED DISTINCTIONS BETWEEN ACHIEVEMENTS

Teachers are well placed to make fine distinctions between the achievements of students whose work they have seen. The process of making comparisons against abstract statements of standards is much less precise – a key reason that usual practice for standards-referenced systems is to have fewer than ten levels<sup>13</sup>.

In principle, and experience suggests in practice, they are likely to be better at decisions about order (who did better than whom) than about gaps (who did better than whom by how much).

Recent developments in techniques and the technologies that support paired comparisons<sup>14</sup> suggest that ideas about making order and gap judgments that were first explored with Queensland teachers in the mid-1980s can be more readily operationalised than proved possible at that time.

---

<sup>13</sup> Queensland currently uses five – levels of achievement are standards-referenced where achievement is matched against explicit statements. New South Wales uses six – bands 1 to 6. South Australia uses five with a plus/minus distinction to get fifteen grades.

<sup>14</sup> See Whitehouse, C. and Pollitt, A (2012). *Using adaptive comparative judgement to obtain a highly reliable rank order in summative assessment*. Manchester: AQA Centre for Education Research and Policy.

## 5.2 THE SCHOOL-DECISION MODEL PUTS THE IMPORTANT DECISIONS WHERE THEY SHOULD BE

Having the school as the decision-maker is fundamental to the current Queensland system of externally-moderated school-based assessment, of which the OP system is a part. This is a strength of the system in the sense that it puts the institution best placed to know and understand a student's achievement right at the centre, rather than using the snap-shot, limited one-off sample approach of the classic external examination, which is often assumed to be the only way in which high-stakes decisions can be made at the senior secondary level<sup>15</sup>

It may be worth remembering that the Queensland system had its origins in the 1970s in community and teacher realisation of the strengths of having schools make decisions about the achievements of their students.

On the other hand, the Queensland system avoids the problems that occur when schools are the sole decision-maker with no external moderating influence that is designed and implemented to ensure that these decisions are soundly based.

School-based assessment alone is well-known to lead to grade inflation and other undesirable consequences for both reliability and validity of results. Any system that bases high stakes decisions on provider-based assessments requires careful design to minimise risks to validity of results, including, as examples,

- pressure from parents make it hard to refuse to give higher results in individual cases
- school reputation pressures push standards downwards in practice
- the negative consequences to the school of not having appropriate practices in these assessments are less than the costs to it of doing so
- the school, does not ensure that its staff implement its teaching and assessment strategies correctly and consistently
- the school gives greater priority to meeting the needs of the learner for the issue of the qualification than to applying the standards

A similar set of issues about validity and reliability of assessment decisions is found in the current Australian VET system, where each provider is responsible for assessment and issue of qualifications. The issues, causes and proposed changes were explored in some depth in a report to the then National Quality Council<sup>16</sup>. One of the key recommendations in this report has emerged in proposals from the National Skills Standards Council<sup>17</sup>. These propose that each VET provider should be required to have an 'Accountable Education Officer'. Such an officer must be appropriately qualified and be registered.

---

<sup>15</sup> Although high-stakes assessment decisions are successfully made in other ways in other areas.

<sup>16</sup> National Strategic Industry Audit TAA40104 Certificate IV in Training and Assessment Stage 2 Report. National Quality Council 2011. Downloaded from [http://www.nssc.natlese.gov.au/\\_data/assets/pdf\\_file/0017/58103/Stage\\_2\\_National\\_Strategic\\_Industry\\_Audit\\_TAA40104\\_20110726.pdf](http://www.nssc.natlese.gov.au/_data/assets/pdf_file/0017/58103/Stage_2_National_Strategic_Industry_Audit_TAA40104_20110726.pdf)

<sup>17</sup> National Skills Standards Council March 2013 *Improving Vocational Education and Training. A position paper of the national skills standards council*. COAG Consultation Regulation Impact Statement. Downloaded from [http://www.nssc.natlese.gov.au/\\_data/assets/pdf\\_file/0020/74063/NSSC\\_Position\\_Paper.pdf](http://www.nssc.natlese.gov.au/_data/assets/pdf_file/0020/74063/NSSC_Position_Paper.pdf)

This registration can be withdrawn where there is a demonstrated failure of assessment practices by the provider.

This proposal is an example of designing a system so that there is a balance of pressures – in this case, the provider’s interest in maximising the number of qualifications (a market pressure that operates at an organisation level) and the accountable officer’s interest in maintaining registration and hence the officer’s role with the provider (a market pressure that operates on an individual level).

The design of the current OP system has a set of elements that provide for a balance of the pressures schools experience in their decision-making about students’ achievements. The successful implementation of this design depends on resources, leadership and commitment.

### 5.3 TRIANGULATION: SAIs, LOAs AND THE QCS TEST

A strength of the current OP system is that it draws on three kinds of data:

- SAIs – the interval scale representing relative achievement within a subject within a school
- Levels of Achievement (LoAs)– externally moderated, standards-based subject results, comparable across the state
- QCS test scores – a reference test of common (not subject-specific) curriculum elements.

This allows for triangulation:

- Levels of Achievement can be used to check for anomalies in OPs – students with the same set of levels of achievement in the same subjects should not have very different OPs
- within-school estimators based solely on SAIs can be used to reduce the influence of any anomalous QCS scores (students who done well at school but poorly in the QCS test and vice-versa)
- measures based on QCS scores align estimates of overall achievement (derived from SAIs) between schools.

Some external examination systems have a form of triangulation in using teacher estimates of students’ predicted results as a way of identifying potentially anomalous results.

### 5.4 ALLOWS WIDE RANGE OF SUBJECT COMBINATIONS FROM A WIDE RANGE OF SUBJECTS

The current OP system allows students to follow a wide variety of subject combinations – in the 1990s, the number of unique subject combinations was roughly one third of the number of students. At the same time, most students include some English and some mathematics in their studies.

However, it should be noted that a comprehensive review in England, *A Review of Vocational Education - the Wolf Report*<sup>18</sup> argues for the value of each student having an overall study program that is

---

<sup>18</sup> Wolf, A. 2011. *A Review of Vocational Education – the Wolf Report*. Downloaded from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/180504/DFE-00031-2011.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf)



governed by a set of principles relating to content, structure, assessment and contact time and that is coherent. The analysis and conclusions of the Wolf Report indicate that wide diversity of courses is not in itself sufficient.

## 5.5 SOPHISTICATED PROCESSES

The current system is underpinned by a wide repertoire of sophisticated quantitative and qualitative processes.

## 6 WEAKNESSES OF THE CURRENT OP SYSTEM:

---

The current system has two types of weaknesses: in principle weaknesses, attributable to key assumptions of the system and in practice weaknesses – ones attributable to changes in external circumstances and or internal actions (or inactions).

### 6.1 IN PRINCIPLE WEAKNESSES:

#### 6.1.1 The OP is focused on university entrance

The OP system's focus on university entrance as its fundamental purpose is an in-principle weakness. It actually provides an estimate of overall achievement in senior secondary studies – something that could be used for a variety of purposes. However, its focus on university entrance creates difficulties and complications including:

- who is eligible and who is not eligible – definitions and implementation
- impacts on patterns of participation in subjects and on perceptions of the purpose of particular subjects
- perceptions that some schools advise students on their study patterns (and whether or not to sit the QCS Test) in terms of the possible impact on school status as reflected through OP distributions
- which types of subject results will be counted – why not anything that is a valid study for senior secondary students
- the use of OP distributions as an indicator of 'school performance' – reinforcing a perception that university entrance is the only important aspect of senior secondary studies
- a focus on the QCS Test as a scaling instrument for university entrance rather than as something providing individual and group information about learning.

#### 6.1.2 Breakdown of fundamental assumptions

1. Assumption 4.1

It is likely that young people will increasingly seek to move through post-year 10 education and training in varied ways, entering and leaving at various times, mixing full and part-time participation with employment and other activities. The current OP system is not built for this.

## 2. Assumption 4.2

The system is not designed to ensure that there is a reasonable range of participation within each school and each subject within the school. It assumes that this will happen (which historically it did) but does not ensure it. The assumption of a reasonable range of participation and achievement is fundamental to a variety of assessment systems – for example, James Popham has pointed out that standardised tests will delete items that all students get right, so if teachers successfully taught all students the required knowledge to the required standards, the types of tests currently used in the US would ensure that this success could not be reflected in test scores.

Only subjects where there is some differentiation of achievement can be used as input. This presents an obstacle to using subjects where results do not permit any real differentiation – for example, so-called competency-based assessment presents results that are at best dichotomous (competent/not yet competent) and usually virtually or completely single-valued – everybody with a result has the same result. Many scaling models, including those in common use such as Item Response Theory, breakdown if the input data is not sufficiently differentiated.

## 3. Assumption 4.3 and 4.4

Effective ‘manipulation’ of SAIs to advantage some students while spreading the consequences across other students is difficult and can be identified and corrected by the QSA. Fundamentally, however, the capacity of schools to make good decisions – a strength of the system – must bring with it a capacity to make poor decisions. A system that values the professionalism of teachers has that as its strength and its weakness. A system that is not based on valuing the professionalism of teachers has that as its weakness – people tend on average to behave in line with such expectations – and its strength – it is ‘teacher-proof’.

As noted in Allen (2012)

*From the outset, the Queensland system located not only responsibility for high-stakes assessment decisions at the individual school level, but responsibility for the details of the course and assessment program that students at this school would follow. This local responsibility followed from the assumption of the professionalism of the teacher, from the idea that learning is best achieved where the teacher is actively designing the learning program, implementing it, designing how information about student achievement will be gathered and gathering and using that information.*

*Positioning of the teacher as a fair and reasonable, professional arbiter, applying standards to students’ work, students with whom the teacher engages on a daily basis, places additional responsibility on the teacher. And at the same time it prevents the teacher from taking the traditional role in an external examination system of being in*

*partnership with the student, seeking the best advantage in a contest with the examiner...*

*The value of having an assessment system based around the ideal of teacher professionalism should not be underestimated: ensuring that classroom practices foster the development of the deep learning considered essential for students' futures requires professional teachers. Systems that espouse one view of teachers but imply another in the way they act, systems that behave as if most teachers cannot be professional will find that many will live down to this expectation – though there will be honourable exceptions. Systems designed around the expectation of professional behaviour will find, over time, that many, though not all, will live up to this expectation.*

## 6.2 IN PRACTICE WEAKNESSES

### 6.2.1 The competitive academic curriculum

The set of subjects with results that count towards the OP is dominated by what Connell and Ashenden called the 'competitive academic curriculum' (CAC). This is marked<sup>19</sup> by characteristics such as

- division of knowledge into 'subjects'
- a hierarchy of subjects – with mathematics amongst the top
- a hierarchical ordering of knowledge within each subject
- formal competitive assessment (the 'exam').

This is not a necessary feature of the system, but the taken-for-granted nature of the CAC – the way the system is built around the assumption of most people that this is the way things are meant to be – means that rigorous learning that doesn't fit the CAC model<sup>20</sup> does not really have a place in the current OP system.

### 6.2.2 Changing patterns of participation<sup>21</sup>

The following graph charts trends over the period 1992 to 2012, each trend as an index with the 1992 value as 100. The trends shown are changes in

- the *total count* of year 12 students
- the *OP count* of students eligible for an OP

---

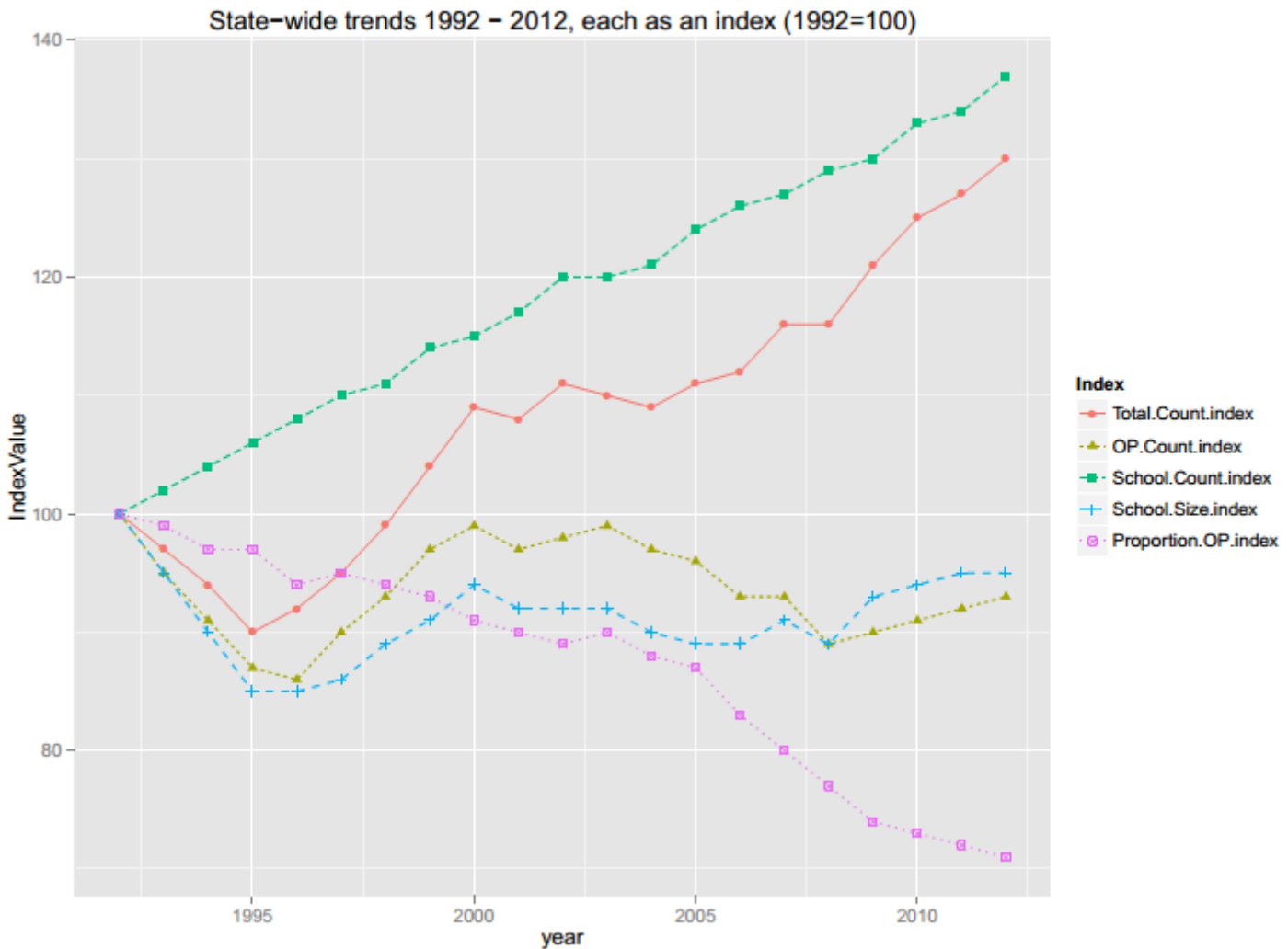
<sup>19</sup> Taken from RW Connell "Social Change and Curriculum Futures" in *Change: Transformations in Education*, vol 1, No. 1 May 1998 pp 84 -90

<sup>20</sup> For example, one based on students' following a single, complete program of what some people call applied or situated learning, one not organised into five or six subjects.

<sup>21</sup> Data sets provided by the Queensland Studies Authority.

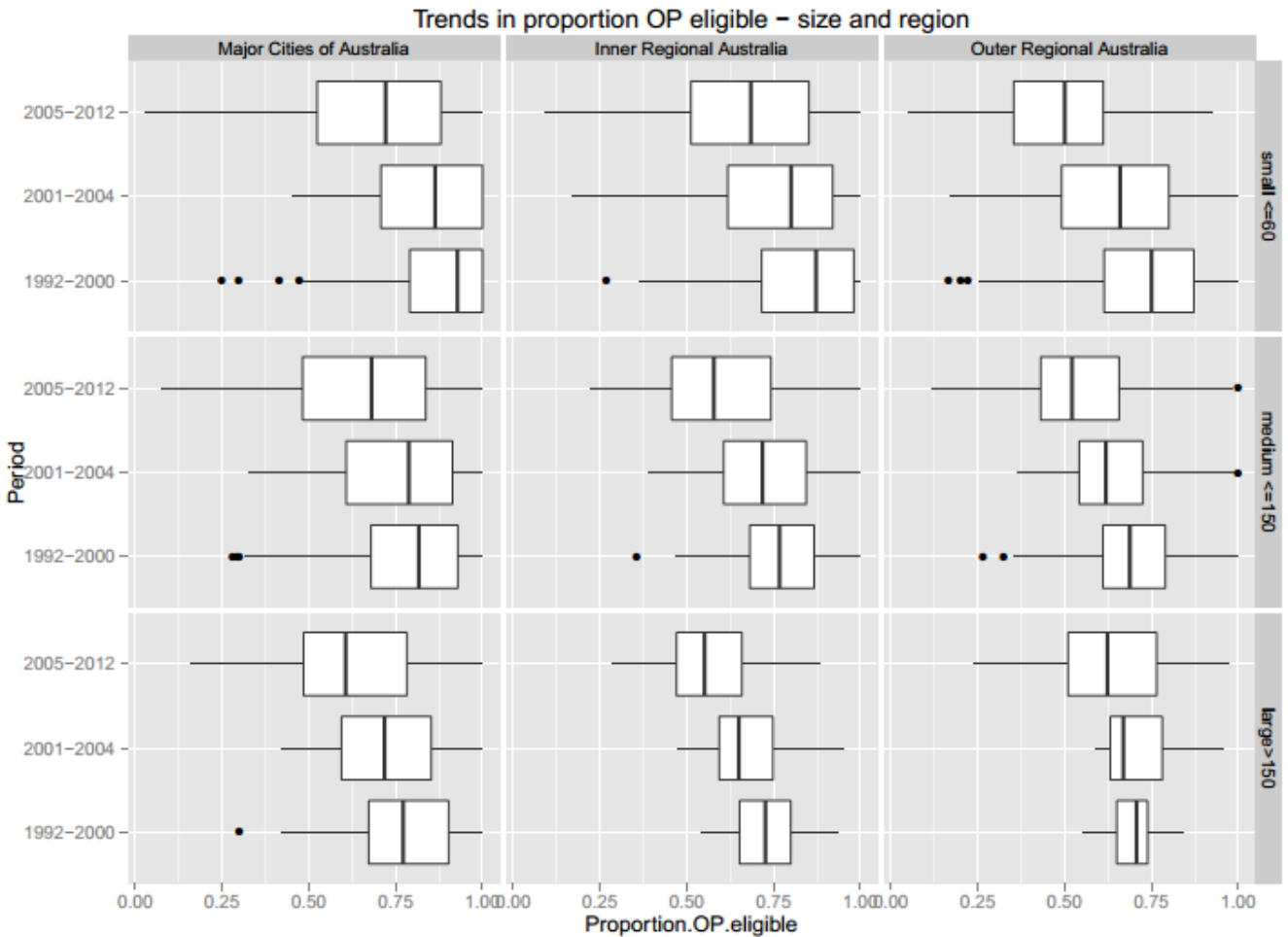
- the *school count* showing the number of schools with year 12 students
- the *school size index* showing the average size (total year 12 count) of schools
- the *proportion OP index* showing the number of OP eligible students as a proportion of the total year 12 count.

Clearly, the number of year 12 students and the number of schools has steadily increased since the year 2000. The number of OP eligible students as a proportion of the total count has steadily declined – despite a slight increase in the absolute numbers since 2008. Since the year 2000, the size of the average school has varied between 89 and 95 per cent of its value in 1992.

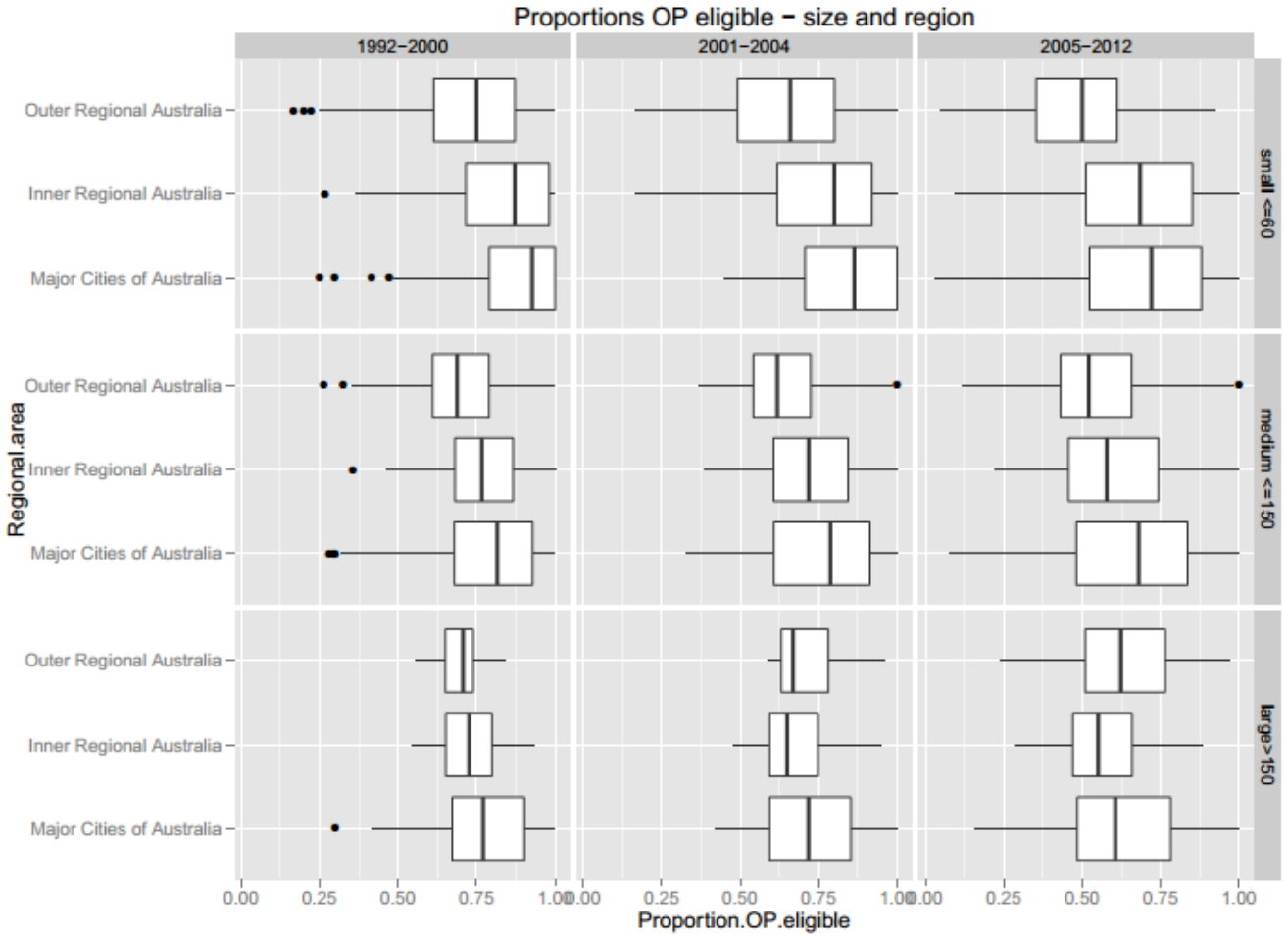


At an individual school level, the trend towards a lower proportion of OP eligible students is evident over time. As is clear from the following graph, there are differences in terms of the size of school (total number of year 12s) : smaller schools are more variable and the range of differences between larger schools of a similar size is increasing.

A brief look at additional data about schools provided by the QSA (location, an index of relative socio-economic disadvantage) suggests that there is considerable complexity in the factors associated with the proportion of OP eligible students<sup>22</sup>. Aspects of the complex relationship are illustrated in the following boxplots.

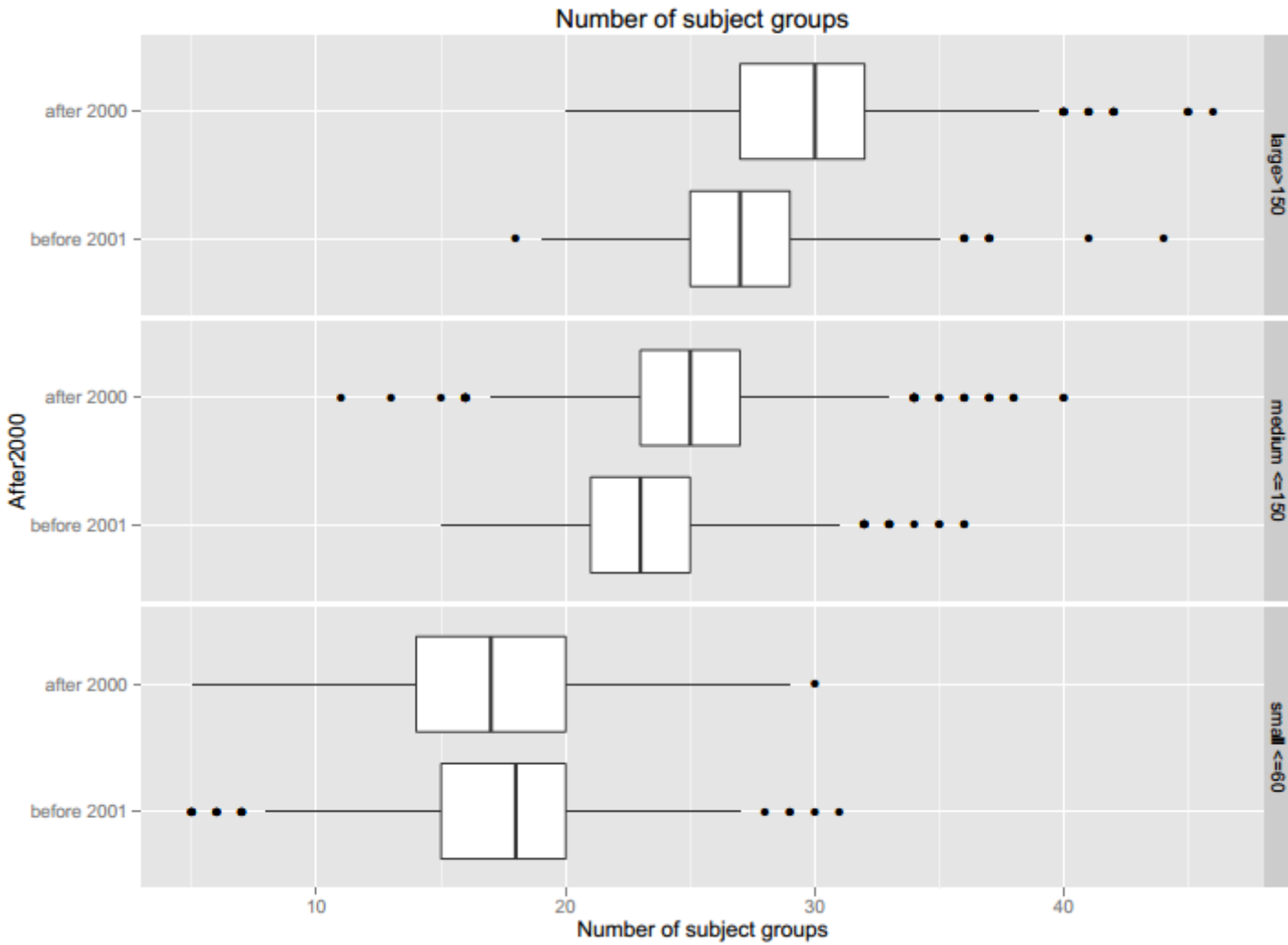


<sup>22</sup> Modelling using general linear modelling (glm) in R: quasibinomial family. There are significant main effects for year, size, Index of disadvantage (IRSD), Regional area and interaction effects of size and IRSD, size and region, region and IRSD.

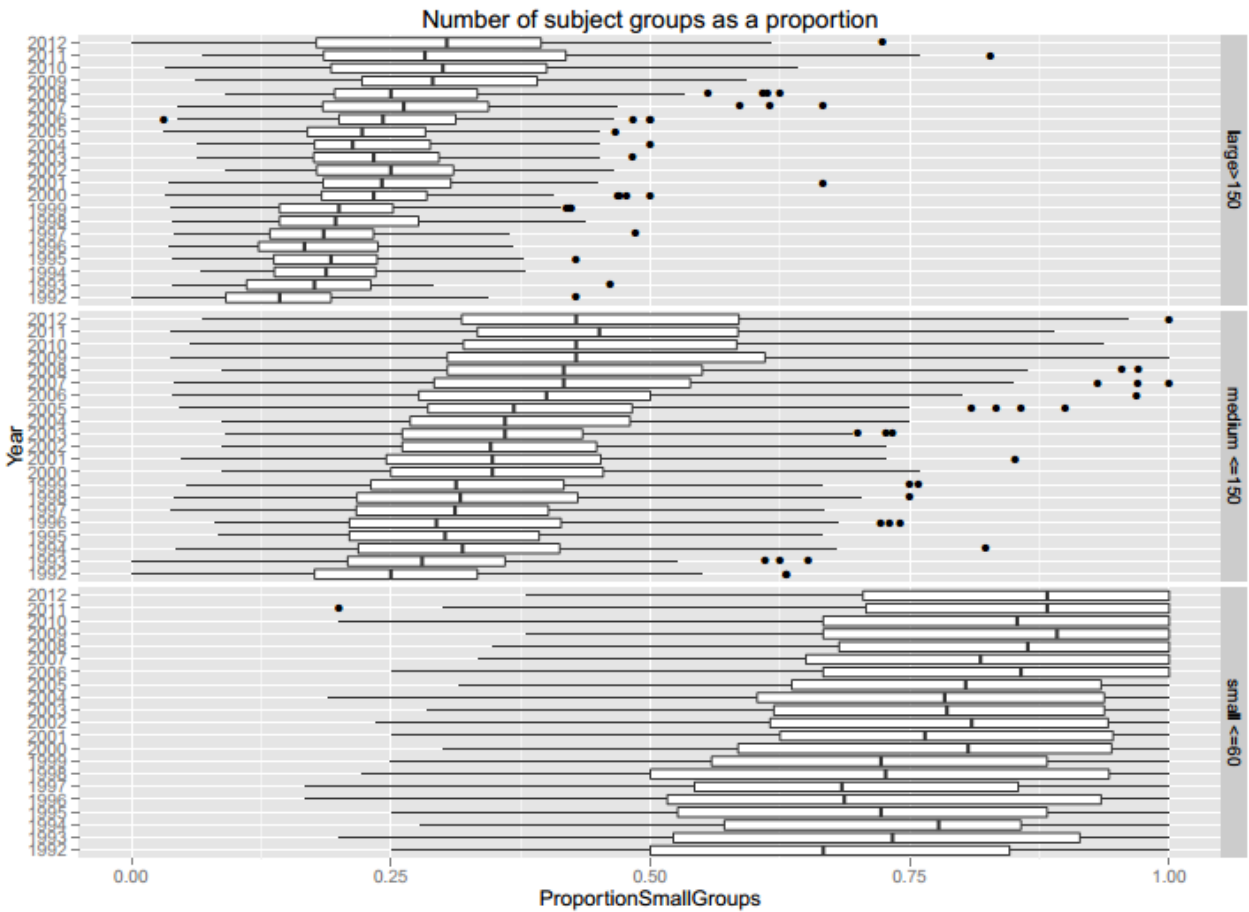


None of these complexities is surprising, of course. Their significance is that observations about the strengths and weaknesses of the current system (and any proposals for change) should be understood in the context that participation rates vary across schools to the extent that discussion based on simple stereotypes about the system and schools is very likely to be misplaced.

The number of subject groups per school is obviously affected by the size of the school, although as noted above there are small schools with a lot of subject groups and large schools with relatively few. Once this basic relationship is accounted for, there are significant trends in recent years towards an increase in the number of groups (again there are complex relationships with other factors) as is evident in the following boxplot.



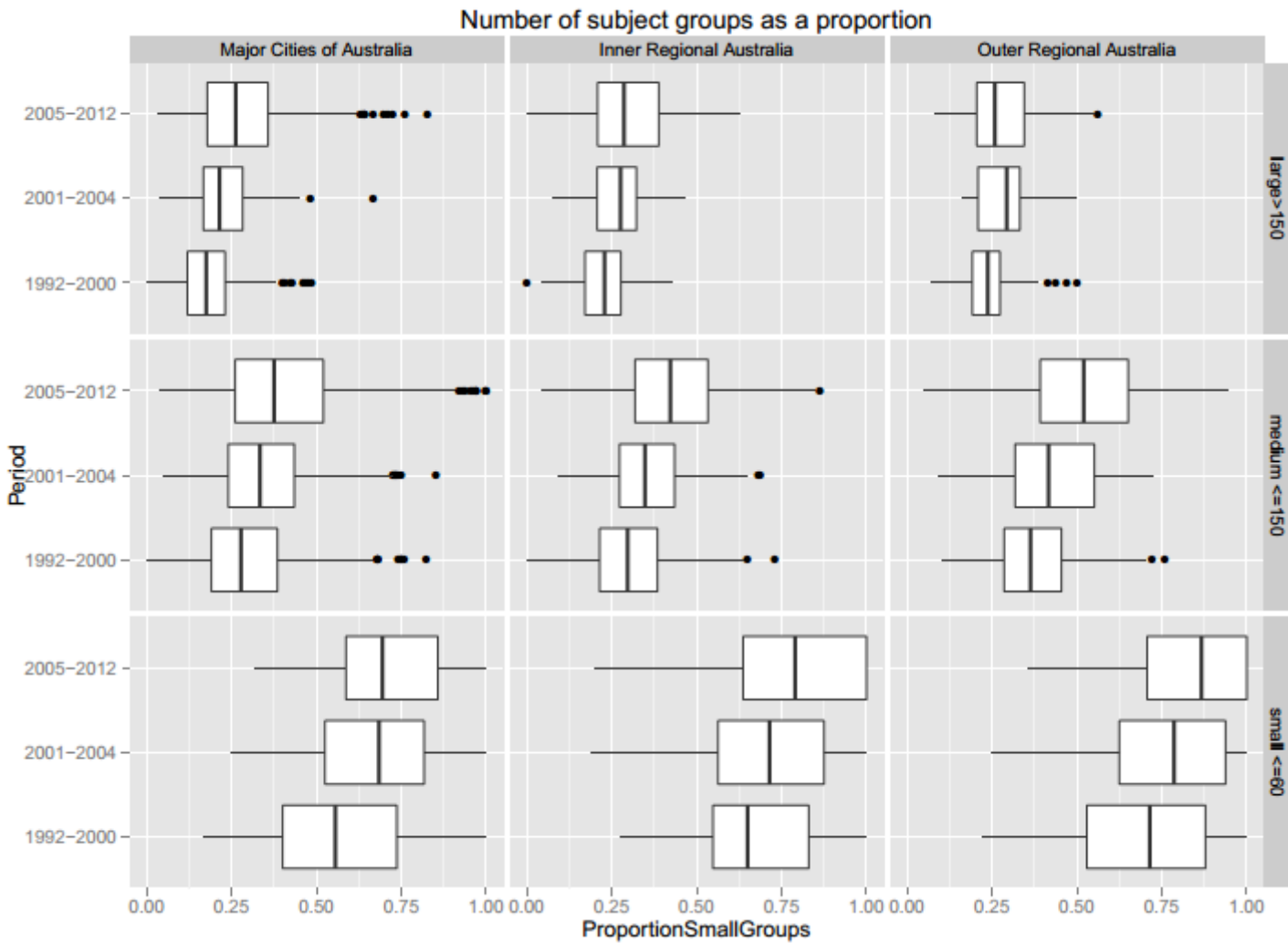
The pressures on the OP system can be gauged from the changes in the proportion of subject groups that are classified as 'small' subject groups<sup>23</sup>. There is a significant increase in the proportion of 'small' subject groups over time.



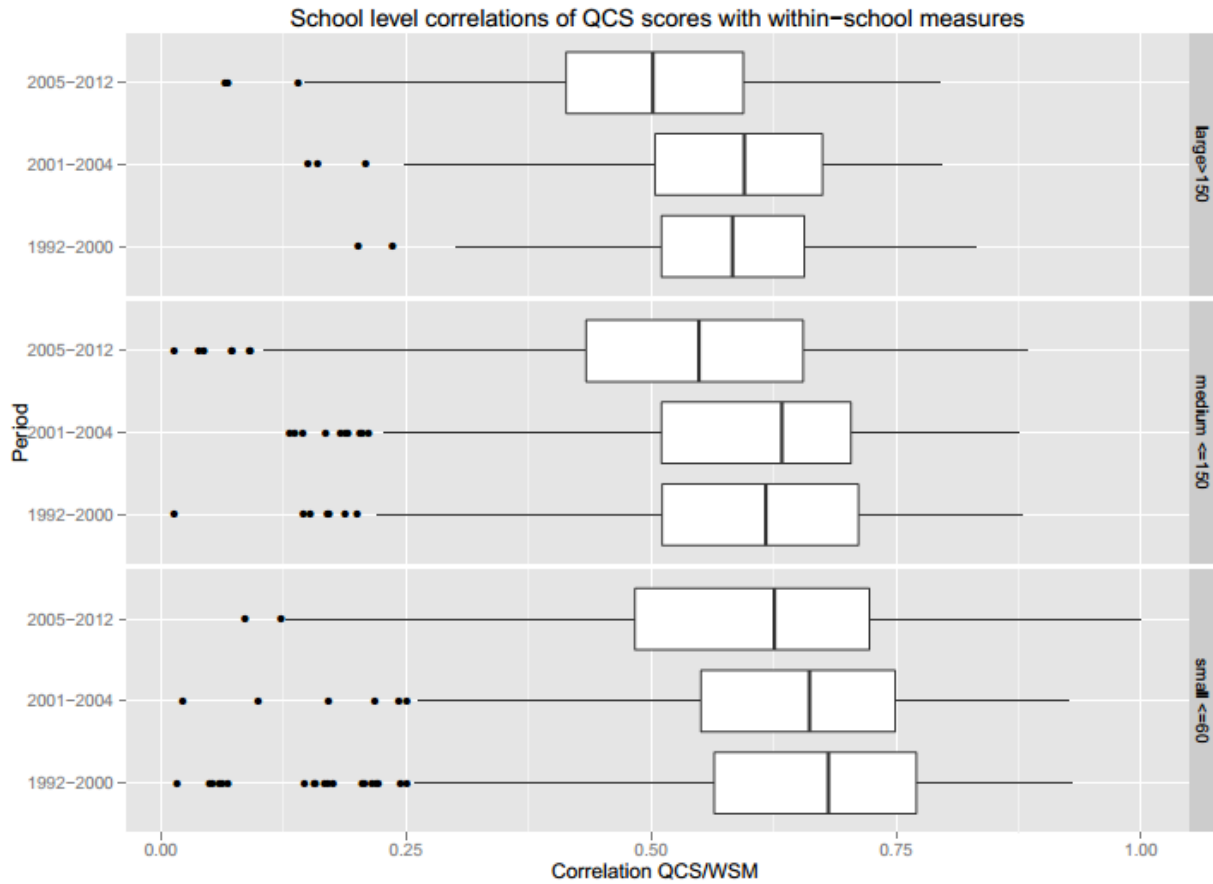
<sup>23</sup> Subject groups are classified as small, intermediate and large. Results in large subject groups are scaled. Results in small subject groups are not (the scaling process makes no sense for a small group of one, although in the early days of the TE score this is what happened). Results in intermediate groups are a weighted average of a small group result and the result of scaling using the procedure for large groups.



There are, as before, complex relationships with other factors. The following boxplot illustrates how the changes over time play out differently in different regions.

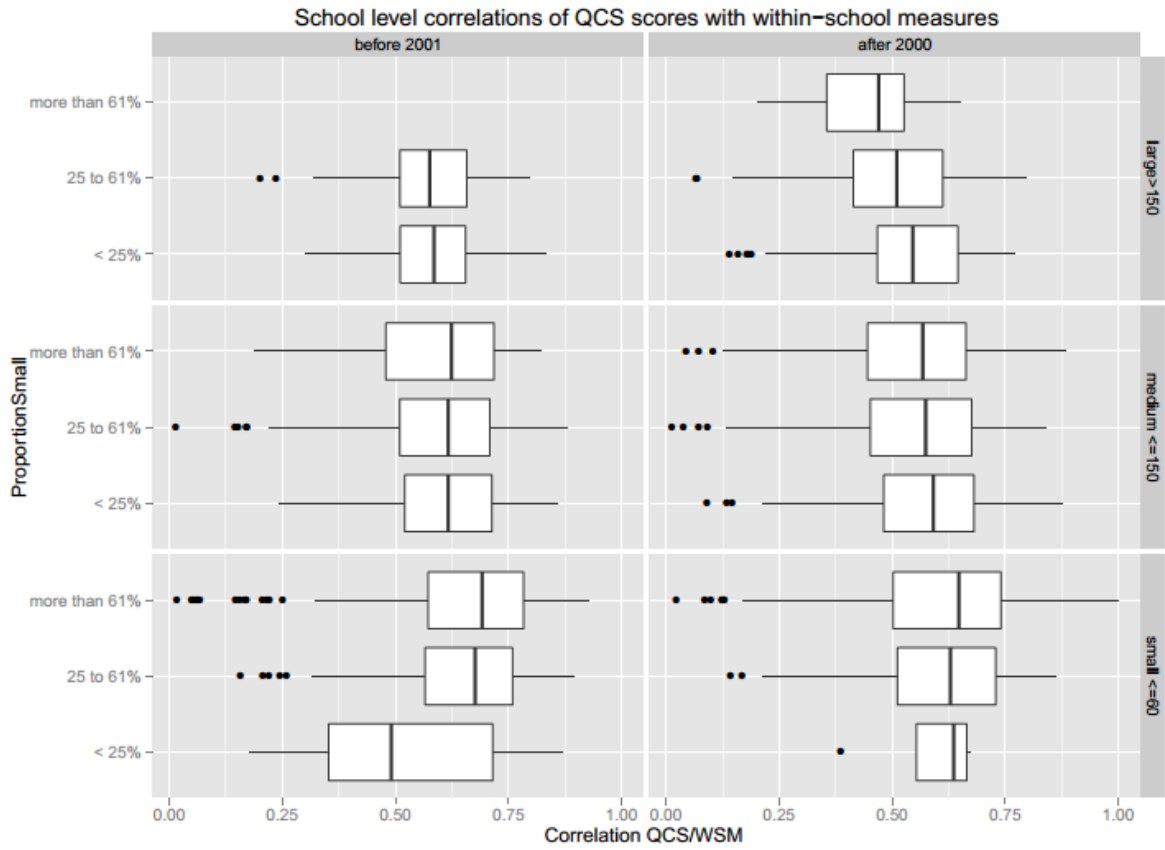


The smaller proportion eligible and the large proportion of small groups are likely to be associated with a reduced range of participation and/or less coherence across subjects<sup>24</sup>. We can get some indication of this by looking at the change in the relationship of QCS scores and within-school estimates of overall achievement.

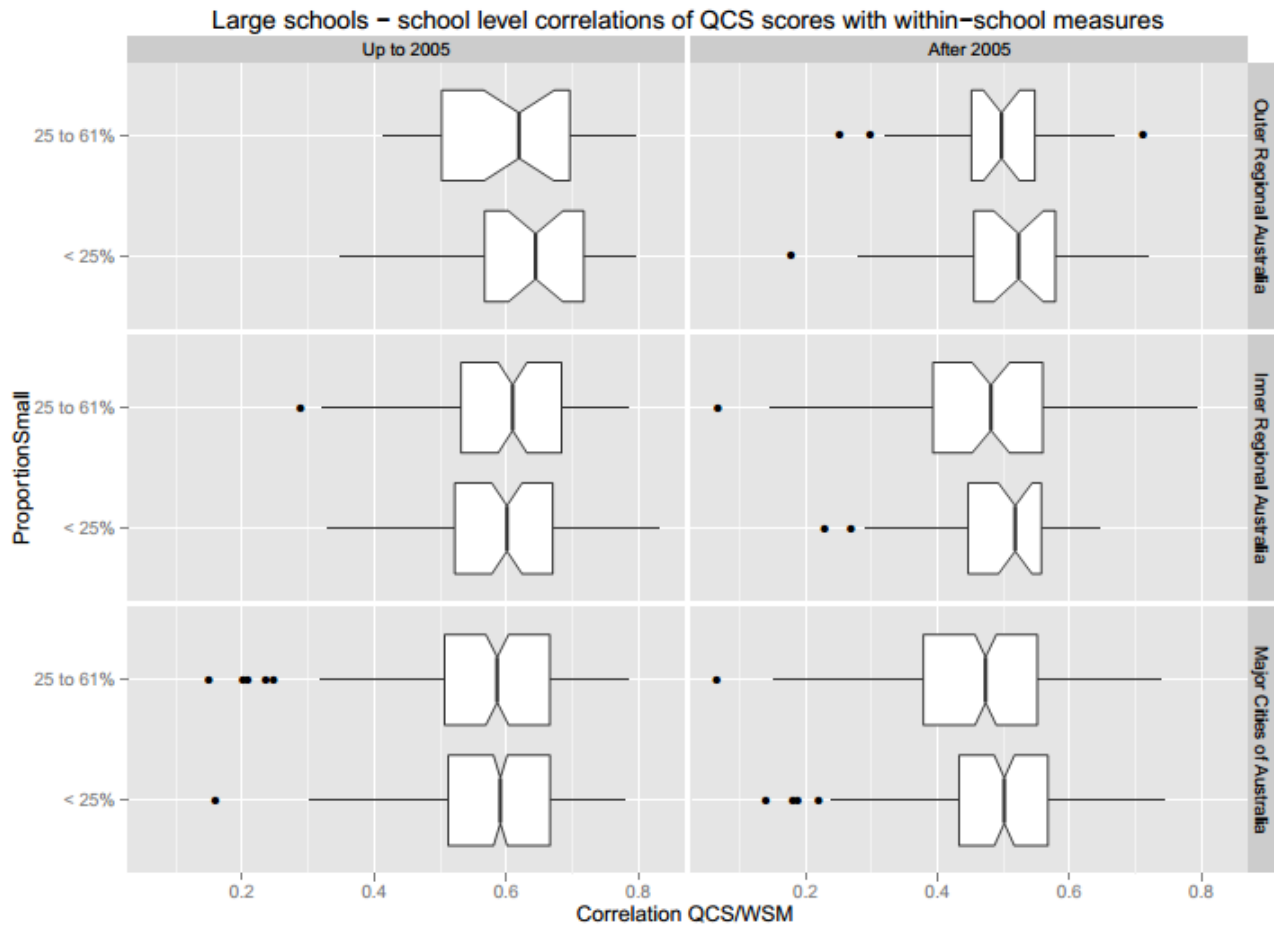


<sup>24</sup> Reduced coherence across subjects can be, but does not have to be, a consequence of reduced range of participation – the more restricted the range of participation the less coherence we should expect on average. However, it is also possible to have less coherence across subjects for other reasons.

There is a noticeable decline in this relationship – consistent with the idea that the participation pattern is becoming more select. This appears to be more a feature of schools with more than 150 year 12 students.



And it seems to have a regional element to it as well, as is suggested by the following boxplot, which is restricted to large schools. The notch in the boxplots suggest confidence intervals for the medians – if the notches do not overlap it is likely that the medians are different.



Up to 2005, the differences between large schools in the metropolitan area with few small groups (less than 25 per cent) and more small groups is not noticeable. After 2005, it is. A similar difference is emerging for large schools in other areas.

These suggest that the changes anticipated in 2001, ones that have consequences for the viability of the system, are underway.

School QCS means and standard deviations in principle might provide further insight, if we could assume that these parameters are on a comparable scale from one year to the next. If this is a poor assumption, we will not gain any understanding of whether participation is becoming more restricted in range.

To look at this we have used linear modelling of each of these parameters from

- the category of school size (small, medium and large)
- the proportion OP eligible
- the proportion of small groups
- the number of subject groups

- the correlation of QCS and Within School Measure
- the regional area
- an index of relative social disadvantage
- the year (as a count of the number of years since 1992).

The data set was restricted to those schools for whom we have measures on all these data elements, where the size of the school was at least 21 and where the school QCS standard deviation was not less than 10 and not more than 80.

The results are shown in the following tables

- Call: lm(formula = School.QCS.mean ~ Size + Proportion.OP.eligible + ProportionSmallGroups + NumberSubjectGroups + QCS.WSM.correlation + Regional.area + scale(IRSD) + I(year - 1992), data = subset(qld.2, Year12Count > 20 & !is.na(Regional.area) & School.QCS.sd < 80 & School.QCS.sd > 10))

- Residuals

Min	1Q	Median	3Q	Max
-37.025	-4.216	0.055	4.100	28.599

- Residuals standard error: 6.6005 on 6240 degrees of freedom

- Multiple R-Squared: **0.4083**

- Adjusted R-Squared: **0.4071**

- F-statistics: **358.7707** on 12 and 6240 DF. P-value: **0**.

- Coefficients

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	146.95020	1.03202	142.391	< 2e-16	***
Sizemedium <=150	1.98984	0.34684	5.737	1.01e-08	***
Sizelarge >150	3.43760	0.49348	6.966	3.60e-12	***
Proportion.OP.eligible	19.29191	0.70935	27.197	< 2e-16	***
ProportionSmallGroups	-3.02562	0.76746	-3.942	8.16e-05	***
NumberSubjectGroups	-0.07797	0.02605	-2.994	0.00277	**
QCS.WSM.correlation	14.36252	0.59006	24.341	< 2e-16	***
Regional.areaInner Regional Australia	3.58896	0.22095	16.243	< 2e-16	***
Regional.areaOuter Regional Australia	2.37508	0.24945	9.521	< 2e-16	***
Regional.areaRemote Australia	2.02124	0.88090	2.295	0.02179	*
Regional.areaVery Remote Australia	-10.03085	1.73131	-5.794	7.22e-09	***
scale(IRSD)	2.34528	0.11069	21.188	< 2e-16	***
I(year - 1992)	0.19115	0.01617	11.818	< 2e-16	***

--- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table

Response: School.QCS.mean

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Size	2	4855	2427	55.7143	< 2.2e-16 ***
Proportion.OP.eligible	1	125866	125866	2889.0619	< 2.2e-16 ***
ProportionSmallGroups	1	332	332	7.6157	0.005803 **
NumberSubjectGroups	1	61	61	1.4047	0.235983
QCS.WSM.correlation	1	20745	20745	476.1606	< 2.2e-16 ***
Regional.area	4	5651	1413	32.4283	< 2.2e-16 ***
scale(IRSD)	1	23970	23970	550.1936	< 2.2e-16 ***
l(year - 1992)	1	6085	6085	139.6697	< 2.2e-16 ***
Residuals	6240	271854	44		

--- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This analysis suggests that higher QCS means are associated with being a larger school with a higher proportion of OP eligible students, fewer small groups, a higher QCS/WSM correlation, being a school with a high IRSD and being outside major cities and very remote areas. When these associations are taken into account there appears to be small upward trend associated with the number of years since 1992.

A similar analysis of school QCS standard deviations follows.

• Call: lm(formula = School.QCS.sd ~ Size + Proportion.OP.eligible + ProportionSmallGroups + NumberSubjectGroups + QCS.WSM.correlation + Regional.area + scale(IRSD) + l(year - 1992), data = subset(qld.2, Year12Count > 20 & !is.na(Regional.area) & School.QCS.sd < 80 & School.QCS.sd > 10))

• Residuals

Min	1Q	Median	3Q	Max
-28.14	-3.86	-0.51	3.27	41.58

• Coefficients

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	37.38813	0.98522	37.949	< 2e-16	***
Sizemedium <=150	0.57343	0.33111	1.732	0.083356	.
Sizelarge >150	0.48774	0.47111	1.035	0.300564	
Proportion.OP.eligible	9.05543	0.67719	13.372	< 2e-16	***
ProportionSmallGroups	2.32978	0.73266	3.180	0.001480	**
NumberSubjectGroups	0.09215	0.02487	3.706	0.000213	***
QCS.WSM.correlation	-31.72163	0.56331	-56.313	< 2e-16	***
Regional.areaInner Regional Australia	0.75228	0.21093	3.566	0.000365	***
Regional.areaOuter Regional Australia	0.85498	0.23814	3.590	0.000333	***
Regional.areaRemote Australia	-1.77835	0.84095	-2.115	0.034497	*
Regional.areaVery Remote Australia	4.26024	1.65280	2.578	0.009972	**
scale(IRSD)	-0.02168	0.10567	-0.205	0.837457	
l(year - 1992)	-0.01286	0.01544	-0.833	0.405141	

--- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

- Residuals standard error: 6.3012 on 6240 degrees of freedom
- Multiple R-Squared: **0.3572**
- Adjusted R-Squared: **0.356**
- F-statistics: **288.9998** on 12 and 6240 DF. P-value: **0.**



## Analysis of Variance Table

Response: School.QCS.sd

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Size	2	2737	1368	34.4618	1.305e-15 ***
Proportion.OP.eligible	1	1498	1498	37.7208	8.665e-10 ***
ProportionSmallGroups	1	231	231	5.8163	0.01591 *
NumberSubjectGroups	1	2809	2809	70.7526	< 2.2e-16 ***
QCS.WSM.correlation	1	128947	128947	3247.6371	< 2.2e-16 ***
Regional.area	4	1443	361	9.0832	2.591e-07 ***
scale(IRSD)	1	5	5	0.1205	0.72854
l(year - 1992)	1	28	28	0.6931	0.40514
Residuals	6240	247759	40		

--- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This analysis suggests that higher QCS standard deviations are associated with being a larger school with a higher proportion of OP eligible students, more small groups, a lower QCS/WSM correlation, being a school outside major cities and remote areas. When these associations are taken into account there is no association with appears to be small upward trend associated with the number of years since 1992.

In both cases, if we add the QCS parameter into the model there is a negative association – higher QCS means are associated with lower QCS standard deviations and vice versa.

The associations of higher correlations of individual QCS scores and Within-school estimates with higher QCS means and lower QCS standard deviations invites comment. In theory , all other things being equal, we would expect a higher QCS/WSM correlation to be associated with groups of students who are relatively consistent across the subjects they study – and, it seems possible, therefore mostly develop more across the common curriculum elements tested by the QCS. Conversely, lower QCS/WSM

correlations are consistent with students being inconsistent performers across their school subjects, leading to QCS scores with a lower mean but greater variability. A look at the weighted average QCS means<sup>25</sup> and weighted average QCS standard deviations from 1992 to 2012 shows indeed a tendency to a lower QCS mean and higher QCS standard deviation.

The increase in the number of small groups and the decrease in QCS/WSM correlations, taken together, suggest that a decrease in the range of participation and achievement of eligible students – although the year 12 group as whole has increased.

This is reflected in references to such changes in QSA documentation to changes in participation patterns. For example, *“Due to the QCE and other factors, it is now very common for fewer than half of the rungs on the Form R6 to be covered as OP-eligible students don’t tend to remain in subjects where they are not likely to achieve an SA or higher.”*

### 6.2.3 Decline in participation – alternative means

Firm figures do not appear to be readily available but it appears that schools give considerable prominence to entry ranks determined for students who complete year 12 without being eligible for an OP. This began as a process for exceptional cases but appears to be becoming more common<sup>26</sup>. The more that this route is taken – for whatever reason – the more restricted the range of participation of OP eligible students.

There are also anecdotal reports of schools negotiating direct relationships with particular institutions. To the extent that this happens (if it does) it cuts across a principle of the OP system of providing a means of comparing applicants on the same basis across the whole state.

### 6.2.4 Bonus schemes and FPs

Several universities have instituted bonus schemes – whereby students who complete studies in specific subjects are awarded bonus points. The original intention of FPs was to provide a selection mechanism based on the types of subjects studied – bonus schemes can cut across this design element where it is intended that selection (not recruitment or encouragement to choose a particular institution) is based on the specifics of a student’s subject choices.

### 6.2.5 Pressures to closely align Levels of Achievement and SAIs

Current documentation of QSA end-of-year procedures operationalise an expectation of alignment of SAIs and levels of achievement and gradations within those levels. The more stringent in practice of such alignment the less that it is possible to use Levels of Achievement as a check on SAIs.

### 6.2.6 Public discourse

It is easy to find references in public discourse about the system to ideas such as “OP Score”, references to weighting subjects and a range of claims at variance with the published material explaining the system. Public understanding of and confidence in any system are a continuing challenge – there are new students and parents every year. While a search of the web can find examples of the ‘myths’

---

<sup>25</sup> This is  $\text{sum}(\text{OPeligiblecount} * \text{QCS mean}) / \text{sum}(\text{OPeligiblecount})$

<sup>26</sup> In 2009, 1506 OP ineligible students were offered a place based on their year 12 results the previous year. This rose to 2511 in 2013.

identified by the QSA, there are plentiful examples on student discussion forums of accurate and reasonable comments on the system. The focus on the distribution of OPs (in the absence of meaningful information) as an indicator of school quality fosters concerns about validity.

#### 6.2.7 The need for change and renewal

There do not appear to have been any significant changes or enhancements to the current system for over ten years, although there have been significant changes in patterns of participation in that time.

## 7 CONCLUSIONS AND IMPLICATIONS

---

The OP system is an integral part of Queensland's system of externally moderated school-based assessment. It has significant strengths in design and in practice but also has some significant weaknesses, some of which are a consequence of changes in patterns of student participation. This means that some changes are required, a need that was identified in the Pitman Report in 2001.

Mature systems need periodic review of their principles and their practices. Their replacement brings renewal and responsiveness to changing contexts. The OP system is a mature system.

The development of changes to the current system can be done through identifying feasible<sup>27</sup> combinations of variations on the key elements and key assumptions listed above. Such combinations can then be evaluated against criteria of fairness, appropriate curriculum backwash effects and likely durability in terms of the changes anticipated in senior secondary studies during the next decade. These are matters for a further paper in this series.

---

<sup>27</sup> What will be feasible will also depend on the nature and extent of other changes to the approach to subject assessment and the role of the school.