The Mechanics of the ACT College Scaling System

What are they doing with my scores?

CMA Conference workshop – A CMA initiative.

Ed Staples: 17 August 2013

This handout provides a mathematical explanation of Method of Moments Scaling, and then goes on to briefly summarize how the scaling has been adapted to the OCS scaling regime of the Office of the Board of Senior Secondary Studies in the ACT. Should you require a more detailed explanation of the technical aspects of the scaling, refer to the Edition 1 White paper entitled "The ACT College Tertiary Course Scaling Procedures" June 2013, published by the Canberra Mathematics Association.

Method of Moments Scaling – The Heart of OCS scaling

1. At the heart of the OCS scaling process is a scaling procedure known as Method of Moments scaling. Generally, the Method of Moments is a parameter estimation technique whereby sample moments (for example sample means and standard deviations) are equated with their corresponding population moments. It doesn't have to be applied to the scaling of scores.

2. In order to understand this scaling procedure, we will use a simple example. Imagine a possible scoring of five dance students by 3 expert judges, given in sheet 1.

student	Judge 1	Judge 2	Judge 3
1	2	6	3
2	10	9	6
3	4	6	4
4	8	9	5
5	6	8	4

Sheet 1

3. We can generalise this sheet by denoting the i^{th} student's raw score from the j^{th} Judge as $x_{i,j}$ so for example $x_{2,3} = 6$.

4. The first concept about this scaling procedure that you must understand is the assumption that any linear scaling of the student raw score $x_{i,j}$ is made up of two components – an estimate of an *ability index* that we could call μ_i AND an *error component* that we could call $e_{i,j}$. The index has only one subscript because it belongs to the student. The error term has two subscripts because different judges will have different error terms, even on the same student.

5. Thus the *key equation* in the scaling is given as:

$$b_j.x_{i,j} + a_j = \mu_i + e_{i,j} \qquad (1)$$

6. Hence, for any 15 scores involving 5 dancers and 3 judges, we could create two important tables, referred to as Table 1 and Table 2:

	J1	J2	J3	*Ave*
S1	$b_1.x_{1,1} + a_1$	$b_2.x_{1,2} + a_2$	$b_3.x_{1,3} + a_3$	
S2	$b_1.x_{2,1} + a_1$	$b_2.x_{2,2} + a_2$	$b_3.x_{2,3} + a_3$	
S3	$b_1.x_{3,1} + a_1$	$b_2.x_{3,2} + a_2$	$b_3.x_{3,3} + a_3$	
S4	$b_1.x_{4,1} + a_1$	$b_2.x_{4,2} + a_2$	$b_3.x_{4,3} + a_3$	
S5	$b_1.x_{5,1} + a_1$	$b_2.x_{5,2} + a_2$	$b_3.x_{5,3} + a_3$	
Ave				

Table 1: Linear transformation table

	J1	J2	J3	*Ave*
S1	$\mu_1 + e_{1,1}$	$\mu_1 + e_{1,2}$	$\mu_1 + e_{1,3}$	
S2	$\mu_2 + e_{2,1}$	$\mu_2 + e_{2,2}$	$\mu_2 + e_{2,3}$	
S3	$\mu_3 + e_{3,1}$	$\mu_3 + e_{3,2}$	$\mu_3 + e_{3,3}$	
S4	$\mu_4 + e_{4,1}$	$\mu_4 + e_{4,2}$	$\mu_4 + e_{4,3}$	
S5	$\mu_5 + e_{5,1}$	$\mu_5 + e_{5,2}$	$\mu_5 + e_{5,3}$	
Ave				

Table 2: Ability index and error table

7. The contents of the cells in Table 1 and Table 2 are just different ways of describing the same numbers. Note that in Table 1, the each Judge has a personalized set of scaling parameters b_j running down the columns and in Table 2 each student has a personalized ability index μ_i running across the rows. This is an important observation.

8. Also notice that in both tables, there is an extra row and column to place averages in.

9. Taking the students first, if we average the row scores for the i^{th} student in *both tables* (since the cell contents are equivalent) we arrive at:

$$Ave_i\left[b_j.x_{i,j} + a_j\right] = \mu_i + Ave_i\left[e_{i,j}\right] \qquad (2)$$

10. Taking the Judges second, if we average the column scores for the j^{th} Judge, again in both tables, we arrive at:

$$b_j \cdot Ave_j\left[x_{i,j}\right] + a_j = Ave_j[\mu_i] + Ave_j\left[e_{i,j}\right] \qquad (3)$$

11. Equations (2) and (3) are very important to understand. The subscripts on the average remind us which way the averages were taken. Spend a little time thinking about the sense of these equations. Note for example that equation (3) is really describing 3 equations, one for each judge, and equation (2) is describing 5 equations, one for each student. Note also the rules around averages that have been used to express these equations.

12. The next step is to make a couple of critical assumptions. The two *error averages* in equations (2) and (3) are not the same. One collects error terms along rows, and the other collects them down columns. However, for the purposes of deriving the correct scaling parameters, *we assume that both of these error averages disappear*! In equation (2) for example, if we set the *average* of all the row errors to zero, the cells in the right most column in each table become μ_i . Therefore making the errors vanish will enable us to find the particular scaling constants b_j and a_j required to make that happen. In equation (3), taking out the last term reveal the critical relationship between b_j and a_j . This means that once we determine the three b_j values, the three a_j values can also be found.

13. Now, is it unreasonable to assume that these terms disappear? Not really, given they are uncorrelated random errors occurring in the judging. In the small example that we are using, we might question the assumption, but with lots of judges and students, it is a reasonable assumption to make. So equations (2) and (3) become:

$$Ave_i\left[b_j.x_{i,j} + a_j\right] = \mu_i \qquad (4)$$

$$a_j = Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right] \qquad (5)$$

14. We have reached the final stage of the development of the scaling parameters. We now derive the formula for b_j.

First, set equation (1)

$$b_j.x_{i,j} + a_j = \mu_i + e_{i,j} \qquad (1)$$

Multiply through by μ_i

$$\mu_i\left[b_j.x_{i,j} + a_j\right] = \mu_i{}^2 + \mu_i \cdot e_{i,j}$$

Substitute for a_j from equation (5)

$$\mu_i\left[b_j.x_{i,j} + Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right]\right] = \mu_i{}^2 + \mu_i \cdot e_{i,j}$$

Expand LHS

$$\mu_i\left[Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right]\right] + b_j \cdot x_{i,j} \cdot \mu_i = \mu_i{}^2 + \mu_i \cdot e_{i,j}$$

Take expected values

$$\left[Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right]\right] \cdot Exp[\mu_i] + b_j \cdot Exp\left[x_{i,j} \cdot \mu_i\right] = Exp[\,\mu_i{}^2] + Exp\left[\,\mu_i \cdot e_{i,j}\right]$$

Because μ_i and $e_{i,j}$ are uncorrelated, $Exp\left[\,\mu_i \cdot e_{i,j}\right] = 0$ and so

$$\left[Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right]\right] \cdot Exp[\mu_i] + b_j \cdot Exp\left[x_{i,j} \cdot \mu_i\right] = Exp[\,\mu_i{}^2]$$

Expand and collect the two b_j terms, noting that $Ave_j\left[x_{i,j}\right] = Exp_j\left[x_{i,j}\right]$

$$b_j\left[Exp\left[x_{i,j} \cdot \mu_i\right] - Exp_j\left[x_{i,j}\right] \cdot Exp[\mu_i]\right] = Exp[\,\mu_i{}^2] - \left[Exp[\mu_i]\right]^2$$

And finally, using the definition of covariance and variance,

$$b_j\left[Cov\left(x_{i,j},\mu_i\right)\right] = Var[\mu_i]$$

Yes!

Then

$$b_j = \frac{Var[\mu_i]}{\left[Cov\left(x_{i,j},\mu_i\right)\right]}$$

Using statistical formulae, we can easily show that this is equivalent to:

$$b_j = \frac{Std[\mu_i]}{Std\left[x_{i,j}\right]\left[Correlation\left(x_{i,j},\mu_i\right)\right]}$$

Or more succinctly

$$b_j = \frac{\sigma_{\mu_i}}{\sigma_{x_{i,j}} \cdot r_{x_{i,j},\mu_i}} \qquad (6)$$

15. This value and the value of $a_j = Ave_j[\mu_i] - b_j \cdot Ave_j\left[x_{i,j}\right]$ from equation (5) are the scaling parameters of the method of moments scaling regime. However we still have one important hurdle to overcome.

16. Equation (6) and (5) imply that b_j and a_j can only be determined if we know the very thing we are trying to find - μ_i ! The only way around this is to commence some convergence process by, say, letting the b_j values be 1 and the a_j values be 0. That is, commence with the raw scores, find the μ_i averages, then work out new values of b_j and a_j which in turn lead to new μ_i averages and so on and so forth until the μ_is converge. This describes the scaling procedure completely.

17. To see it all in action, we will use our original data on the dancers and Judges.

18. First let's look back at the original scores of sheet 1

student	Judge 1	Judge 2	Judge 3
1	2	6	3
2	10	9	6
3	4	6	4
4	8	9	5
5	6	8	4

Sheet 1

Setting b_j=1 and $a_j = 0$ for all 3 Judges, we create our starting point

b	1	1	1	
a	0	0	0	
student	Judge 1	Judge 2	Judge 3	Average
1	2.000	6.000	3.000	3.67
2	10.000	9.000	6.000	8.33
3	4.000	6.000	4.000	4.67
4	8.000	9.000	5.000	7.33
5	6.000	8.000	4.000	6.00
Average	**6.0**	**7.6**	**4.4**	**6.0**
SD	**2.8**	**1.4**	**1.0**	**1.7**

Sheet 2

The row averages have been worked out on each student, so we can work on scaling each of the Judges scores using formula (6) and formula (5). The results are shown in sheet 3:

b	0.601852	1.313131	1.733333	
a	2.388889	-3.9798	-1.62667	
student	Judge 1	Judge 2	Judge 3	Average
1	3.593	3.899	3.573	3.69
2	8.407	7.838	8.773	8.34
3	4.796	3.899	5.307	4.67
4	7.204	7.838	7.040	7.36
5	6.000	6.525	5.307	5.94
Average	**6.0**	**6.0**	**6.0**	**6.0**
SD	**1.7**	**1.8**	**1.8**	**1.7**

Sheet 3

The scaling parameters shown after one iteration have made a large impact on the scores. Now, with my little Excel exercise, each new sheet generated provides the *additional* adjustment to the parameters only. So the parameters shown at the top of the sheet will converge toward 1 and 0. In the real world, we would like to know the parameter values of the final scaled scores against the original raw scores. Onto sheet 4 with the new averages!

b	1.000623	1.001291	0.998092	
a	-0.00374	-0.00775	0.011449	
student	Judge 1	Judge 2	Judge 3	Average
1	3.591	3.896	3.578	3.69
2	8.409	7.841	8.768	8.34
3	4.796	3.896	5.308	4.67
4	7.204	7.841	7.038	7.36
5	6.000	6.526	5.308	5.94
Average	**6.0**	**6.0**	**6.0**	**6.0**
SD	**1.7**	**1.8**	**1.8**	**1.7**

Sheet 4

Again sheet 5…

b	1.000011	0.999904	1.000085	
a	-6.7E-05	0.000575	-0.00051	
student	Judge 1	Judge 2	Judge 3	Average
1	3.590	3.894	3.583	3.69
2	8.410	7.843	8.763	8.34
3	4.795	3.894	5.309	4.67
4	7.205	7.843	7.036	7.36
5	6.000	6.527	5.309	5.95
Average	**6.0**	**6.0**	**6.0**	**6.0**
SD	**1.7**	**1.8**	**1.8**	**1.7**

Sheet 5

Sheet 5's parameters are converging to 1 and 0 quickly, so for the purposes of this exercise we'll stop here. From this point, there will be very little movement in the scaled scores. The right hand column shows good estimates on each student's ability index. In Method of Moments scaling it is *these error free ability indices that become the ranking device*.

To finish, Sheet 6 shows the cell differences between the raw scores and the scaled scores of sheet 4. These differences indicate the errors in the original judgments.

b	1.000011	0.999904	1.000085		
a	-6.7E-05	0.000575	-0.00051	*Difference from*	
student	Judge 1	Judge 2	Judge 3	*True Ability*	*True Ability*
1	-1.688	2.312	-0.688	-0.02	**3.69**
2	1.661	0.661	-2.339	-0.01	**8.34**
3	-0.667	1.333	-0.667	0.00	**4.67**
4	0.639	1.639	-2.361	-0.03	**7.36**
5	0.055	2.055	-1.945	0.06	**5.94**
Average	0.00	1.60	-1.60	0.0	
SD	1.1	0.6	0.8	0.0	

Sheet 6

19. What follows is a broad description of how the method of moments scaling technique is adapted to the scaling system used by the Office of the Board of Senior Secondary Studies.

A comparison with the old days

1. In respect of the original ACT College scaling strategy of the last millennium, course scores, irrespective of the college from which they were derived, were simply scaled to the mean and standard deviation of the particular group's AST mean and standard deviation. That is to say, the scaled score $y_{i,j}$ was derived by converting $x_{i,j}$ to a Z score, and then scaled to a mean of $Ave(C_{i,j})$ and standard deviation $Std(C_{i,j})$ where $C_{i,j}$ denotes the corresponding set of AST scores. As a formula:

$$y_{i,j} = \frac{[x_{i,j} - Ave(x_{i,j})]}{Std(x_{i,j})} \cdot Std(C_{i,j}) + Ave(C_{i,j}) \qquad (7)$$

2. This easily understood scaling device was in use for well over a decade, but concern was always expressed about the poor correlations that were occurring between the scaling device (the AST scores) and the course scores. It was this reason that the board began to search around for a better model.
3. However, the Method of Moments result is surprisingly similar. To see this, we can rearrange (7) to make the linear scaling property obvious, so that:

$$y_{i,j} = \frac{Std(C_{i,j})}{Std(x_{i,j})} \cdot x_{i,j} + \left[Ave(C_{i,j}) - \frac{Std(C_{i,j})}{Std(x_{i,j})} \cdot Ave(x_{i,j})\right]$$

4. Compare this with the Method of Moments result when results (5) and (6) are combined.

Then:

$$y_{i,j} = \frac{Std[\mu_i]}{Std[x_{i,j}][Cor\ (x_{i,j}, \mu_i)]} \cdot x_{i,j} + \left[Ave_j(\mu_i) - \frac{Std[\mu_i]}{Std[x_{i,j}][Cor\ (x_{i,j}, \mu_i)]} \cdot Ave(x_{i,j})\right]$$

5. You can see that there are only two real differences. The first is the exchange of the scaling device, and the second is the correlation factor in the b term. Now suppose that there was a perfect correlation between the Judges scores and the ability indices – Think about what would happen.

Developing a model for colleges

1. We now turn to reality, and discuss how the Method of Moments scaling technique is applied to the OCS scaling regime. The detail gets fairly complicated, but the fundamental ideas can be understood readily.

2. There are moderation groups within the participating institutions, and there are around 30 institutions in the system. Groups can be large (greater than 20) or small (less than 11) or in between, but the simple descriptions to follow pertain to large groups. There are all sorts of adjustments to other groups, but they need not concern us here.

Let's now put down the basic procedure.

3. The "judges" scores become the *moderation group scores*, and the "dancers" become the students. The ability indices μ_i , which were simple row averages in the pure Method of Moments (MoM) strategy, change to the scaling scores $V_{i,k}$ which are an average of a student's best 3.6 course scores AND 0.8 of an AST score, in institution k.

4. Not every "student" is assessed by every "judge". That is to say not every student does every course. However, the AST scores themselves become an extra "Judge" (as if the set of AST scores became a course itself). The two important differences between the AST scores, and other group scores in the institution, is that firstly every student in the institution has an AST score, and secondly, that score is a *system-wide moderated score*.

5. So here's what happens. All course scores undergo a preliminary scaling simply to equate the means and standard deviations. Then each moderation group undergoes an MoM scaling using $V_{i,j,k}$ to produce scaling parameters $(b_{j,k}, a_{j,k})$ unique to that group (note the need to identify the college in the subscripts). These linear transformations will ultimately distinguish the groups within the institution, because the only scores compared in any particular group are the n matched pairings $\left(x_{i,j}, V_{i,j,k}\right)$ corresponding to the n students in the moderation group under consideration. This means essentially that a clever group, as shown by $C_{i,j}$, will produce better scaling parameters than one which is not so clever. This part of the process is generally known as *in-school scaling*.

6. In addition to this, the *complete set of AST scores in any one institution*, which we could call $C_{i,k}$ is MoM scaled against the scaling score $V_{i,k}$ to produce a set of *institution scaling parameters* B_k and A_k. These are the *across-school* parameters and they will be used to distinguish the results across different institutions. Theoretically at least, we scale the AST results themselves using B_k and A_k.

7. In practise, the *in-school* and *across-school* scaling features are combined into the one scaling formula, which performs both jobs at once! The "new" b terms are constructed by taking each one of the $b_{j,k}$ s for an institution and *dividing* them by the single institution parameter B_k . We also construct "new" $a_{j,k}$ terms so that in the end the course scaling lines are reflecting both the in-school and across school differences. In the process, we have to *de-scale* the AST scores back to their original form using parameters (1, 0). In simple terms,

we perform the two scaling manoeuvres on the course scores and leave the AST scores alone!

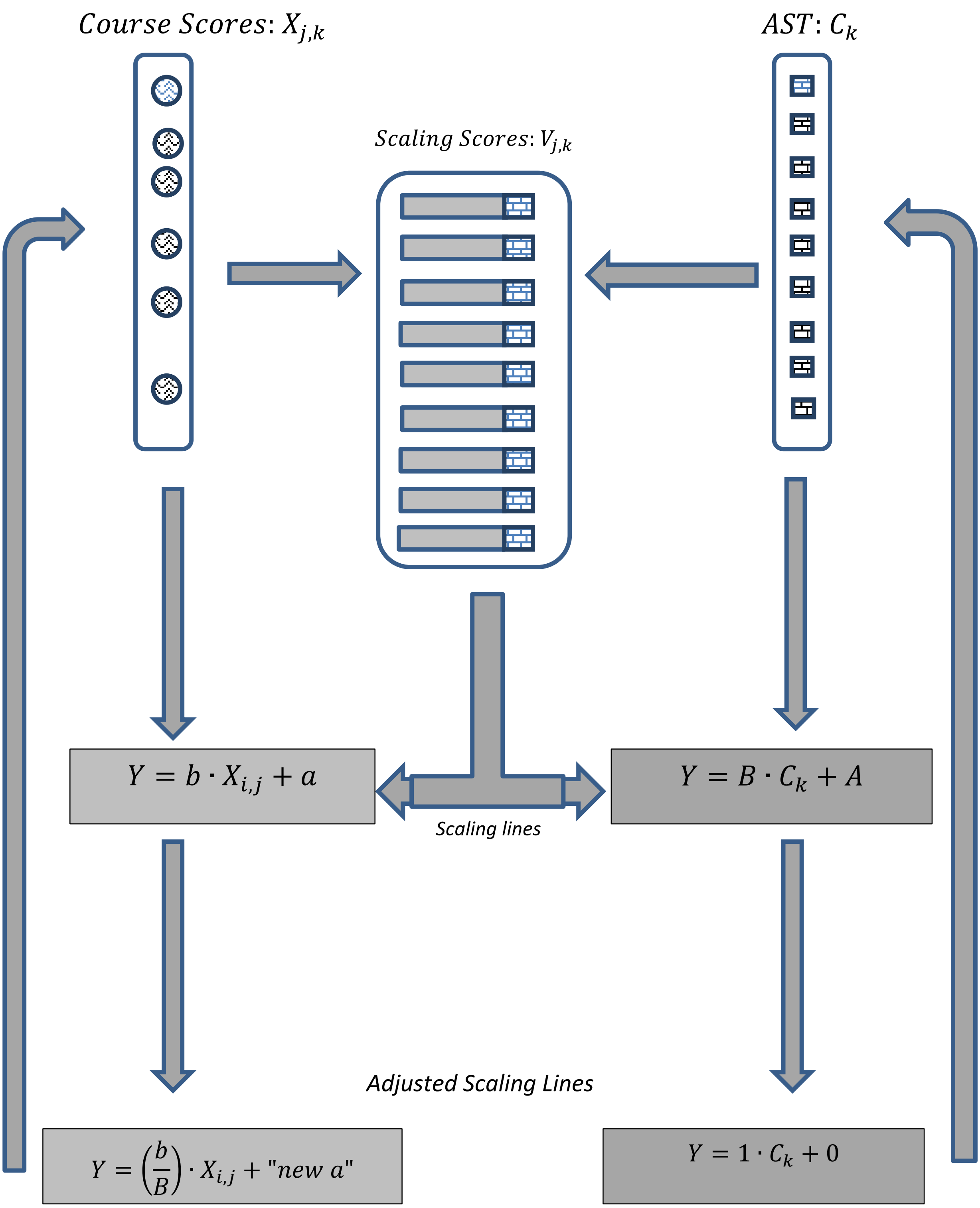
8. The iteration processes that need to occur to find the $V_{i,k}$s is essentially the same as for our simple dance student example. Student averages are formed using the best 3.6 major scores and 0.8 of the AST score. These are our first estimates of the $V_{i,k}$s. From these, the course scaling parameters $(b_{j,k}, a_{j,k})$ and institution parameters (B_k, A_k) are found, and combined in a way that makes the "new" b term $\frac{b_{j,k}}{B_k}$ and the "new" a term given by $\frac{Ave(V_{i,k})-A_k}{B_k} - \frac{b_{j,k}}{B_k} \cdot Ave(x_{i,j,k})$. While the "new" a term looks a little complicated, just think of it as the translation required on the original parameters $(b_{j,k}, a_{j,k})$ to return the AST scaling parameters back to (1, 0).

9. After applying the "new" parameters to the course scores and leaving the AST alone, the second iteration commences. Then the third, the fourth etc until the $V_{i,k}$s converge.

10. At this point two observations can be made. Firstly, because AST remains fixed, it acts as an anchor on the course means. Different course means converge toward their corresponding AST means. Secondly, the 2/11th mix of AST in the average implies that the spread of the course scores would in a small way reflect the spread of the AST scores.

11. Some comment has been expressed as to the effect of the weighting of AST in the $V_{i,k}$s. It is clear that the larger the weighting the more the $V_{i,k}$s will look like the AST scores. Suppose hypothetically we were to consider increasing the weighting of AST in the $V_{i,k}$s ten-fold. Then the parameters (B_k, A_k) would not differ much from (1, 0), and the course parameters $(b_{j,k}, a_{j,k})$ would force the spread of the course scores to be much more like the spread of the AST scores. The lesson is that the higher the mix, the more similar the course spreads will be to the AST spreads on each course.

12. The following page outlines this process diagrammatically.



Schematic diagram of the OCS process