**TERTIARY INSTITUTIONS SERVICE CENTRE**
(Incorporated in Western Australia)

100 Royal Street
East Perth, Western Australia 6004
Telephone (08) 9318 8000
Facsimile (08) 9225 7050
http://www.tisc.edu.au/

# THE AUSTRALIAN TERTIARY ADMISSION RANK (ATAR) - 2017

The purpose of this paper is to explain some of the technical aspects in the calculation of the Australian Tertiary Admission Rank (ATAR) in Western Australia.

ATARs are based on TEAs (Tertiary Entrance Aggregates). In Western Australia, the TEA is the sum of the best four scaled scores (subject to unacceptable combination rules), plus 10% of the best scaled score in a LOTE (Language Other Than English) subject and 10% of scaled scores in Mathematics Methods and Mathematics Specialist. From 2017, the maximum possible TEA is 430.

The TEA provides a ranking within the candidature, that is, the set of students eligible for an ATAR. The ATAR ranks the candidature within the underlying school leaving age cohort. For example, a candidate with an ATAR of 80 is judged to be, in terms of suitability for university admission, at the 80$^{th}$ percentile of the entire school leaving age cohort.

The transformation of ranks based on TEAs to ATARs relies on a participation function. This gives, for each ATAR level, the participation rate, i.e. the proportion of the people in entire school leaving age cohort who are eligible for an ATAR.

For a number of years, most states used a two-parameter logistic function to model the participation function. However, this is no longer feasible because the data used to estimate one of the parameters is no longer available.

In 2015 Harrison and Hyndman proposed the use of a one-parameter family of cubic spline functions as models of the participation function [HH]. The parameter is the ‘overall participation rate’, a weighted average of the proportions of 16, 17, 18, 19 and 20 year olds in the state who are eligible for an ATAR. See the **Note** for the precise definition. The graphs of these cubic splines have sigmoid shapes, and align closely to the logistic curves and the few ‘real’ participation curves that are available. Figure 1 shows the graph of a typical cubic spline participation function.

The proposal by Harrison and Hyndman was endorsed by the Australasian Conference of Tertiary Admissions Centres (ACTAC) for use in 2016 and beyond.

The possible ATAR values range from 99.95 down to 0, in steps of 0.05. The participation function determines an upper bound for the allowed number of ATARs at or above each possible ATAR value. ATARs are assigned to candidates in the order determined by their TEAs, working downwards from the highest. At each stage the highest ATAR consistent with the cumulative upper bounds is assigned.

This procedure produces an ATAR corresponding to each of the TEA values obtained by candidates. To determine a complete mapping of possible TEA values to ATARs, we assign to any possible TEA value $x$, the maximum of all the ATARs that correspond to the TEAs obtained by candidates and are less than or equal to $x$. Figure 2 shows the graph of a typical TEA to ATAR mapping.

## Note

The determination of the Overall Participation Rate (OPR) accords with the procedure prescribed by ACTAC. In WA the candidature of 20 year olds is assumed to be zero, as these students are able to obtain a TEA based on only two courses. The candidature includes onshore international students but excludes offshore international students.

For $i = 16, 17, 18, 19$ and $20$, let

$M_i$ denote the number of eligible candidates of age $i$, and let
$N_i$ denote the number of people of age $i$ in the state (according to ABS data).
Then $M = M_{16} + M_{17} + M_{18} + M_{19} + M_{20}$ is the total number of eligible students.

The Weighted Age Cohort Population $N$ is defined by

$$N = p_{16}N_{16} + p_{17}N_{17} + p_{18}N_{18} + p_{19}N_{19} + p_{20}N_{20}, \text{ where } p_i = M_i/M$$

The Overall Participation Rate (OPR) is defined, as a percentage, by:

$$\text{OPR} = \frac{M}{N} \times 100\%$$

## Reference


[HH] *Modelling the participation function with a one-parameter family of cubic splines*, K.J. Harrison & R.J. Hyndman (2015)


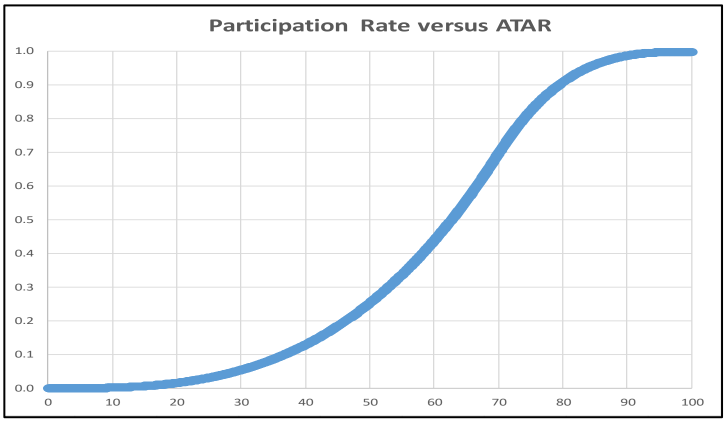


**Figure 1**

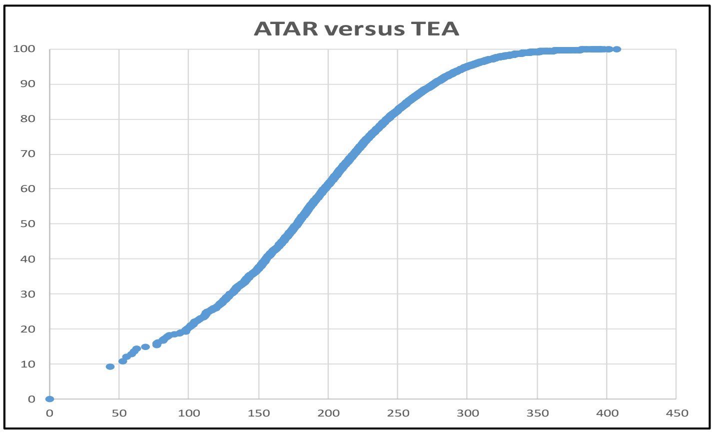


**Figure 2**

# Modelling the participation function with a one-parameter family of cubic splines

Ken J Harrison[1] and Rob J Hyndman[2]

19 May 2015

**Abstract**

We suggest that a simple one-parameter family of cubic spline functions would serve quite adequately as models of the participation curve that is the key component of ITI calculations. This would remove the subjectivity associated with the use of two-parameter logistic functions, and would allow all states to use the same method for ITI calculations.

## 1 Background

The calculation of the Interstate Transfer Index (ITI) involves calculating the "participation curve" in each state, which maps the ranking of a student within the group of students who are eligible for an ITI, to the ranking of the student within the total cohort of students. For the last 18 years, most states have modelled the participation curve using a two-parameter logistic function proposed by Tim Brown.

In 2014, the Australian Conference of Tertiary Admission Centres (ACTAC) decided that it was timely to review the methodology for calculating the Interstate Transfer Index (ITI). The impetus for the review came from several factors. The first was simply that the existing methodology had been in place for many years. A second, more important reason was an unease in some quarters with the subjectivity and lack of uniformity in applying the logistic model for determining the participation curve. A third motiviation was the change in NSW where data that had been used in the ITI calculations were no longer available. In particular, this provided an anchor against which the calculations in the other states could be compared.

Two parameters are required to uniquely specify a logistic curve, and while there is reasonably good agreement about how one of these (the overall participation rate) is defined, there is no such agreement about a second. In NSW the two parameters were determined using logistic regression on data relating to completion rates and the rankings of Year 10 students. Year 10

[1]Murdoch University, WA
[2]Monash University, VIC

data has never been available in other states and is no longer available in NSW. For 2014 results, NSW continued to use logistic regression (NSW Universities Admission Centre, 2014), where the parameters were chosen to match historical data. There was also an adjustment to ensure that there was 100% participation at the highest ITI level.

Professor Tim Brown was appointed to conduct the review, and his report (Brown, 2015) was completed in January 2015. Brown essentially endorsed the status quo, but suggested that suitable data such as common pre-Year 12 assessment should be found and used to validate the choice of participation curves. The only such data currently available is Year 9 NAPLAN results. He indicated that an investigation of the association between Victorian NAPLAN results with Year 12 participation levels and ATARs confirmed that this data was indeed suitable, and he recommended that a wider investigation into the usefulness of NAPLAN results be carried out.

Brown (2015) also rejected a suggestion by the first author (Harrison, 2014) that a one-parameter family of power curves be used as participation curves, despite their advantages of simplicity and transparency. Brown noted that power functions did not give good matches for simulated populations with low participation rates.

Here we recommend the use of a different one-parameter family of functions as participation curves. These are cubic splines and, unlike power functions, their graphs have the sigmoid shape. This is a key feature of logistic curves and the few 'real' participation curves that are available.

Our investigations show that cubic splines provide very good matches to the actual 2014 data in most states. They can be applied without any reference to pre-Year 12 data such as NAPLAN data, and there is no subjectivity in their application as the one parameter can be calculated from the known state participation rate. We believe the simplicity and transparency of this approach resolves the existing problems with ITI calculations.

We describe the model in the following section, and demonstrate its application graphically on 2014 data in Section 3. We summarise our recommendations in Section 4.

## 2 A simple one-parameter participation curve

We propose a new model for the participation curve, $f_p(x)$, where $p$ is the participation rate, $x$ is the proportion ranking[3] of a student within the group of students who are eligible for an ITI, and $f_p(x)$ is the proportion ranking of the student within the total cohort of students. There are no free parameters to be selected or to be estimated from data. The participation rate $p$ determines the curve to be used.

[3]The "proportion ranking" takes values between 0 and 1 and gives the proportion of students who receive a score less than $x$. The usual percentile ranking is simply 100 times the proportion ranking.

The participation curve is defined as follows for $0.25 \leq p \leq 0.75$ (which currently includes all states except Queensland):

$$f_p(x) = \begin{cases} x^3/\alpha^2 & \text{if } 0 \leq x \leq \alpha; \\ 1 - (1-x)^3/(1-\alpha)^2 & \text{if } \alpha \leq x \leq 1; \end{cases} \quad (1)$$

where $\alpha = 1.5 - 2p$.

For $p > 0.75$ (which currently includes only Queensland), the participation curve is defined as

$$f_p(x) = 1 - (1-x)^{p/(1-p)}. \quad (2)$$

For $p < 0.25$ (which is unlikely to occur in any state in the near future), the participation curve is defined as

$$f_p(x) = x^{(1-p)/p}. \quad (3)$$

Figure 1 shows these participation curves for a range of values of $p$.

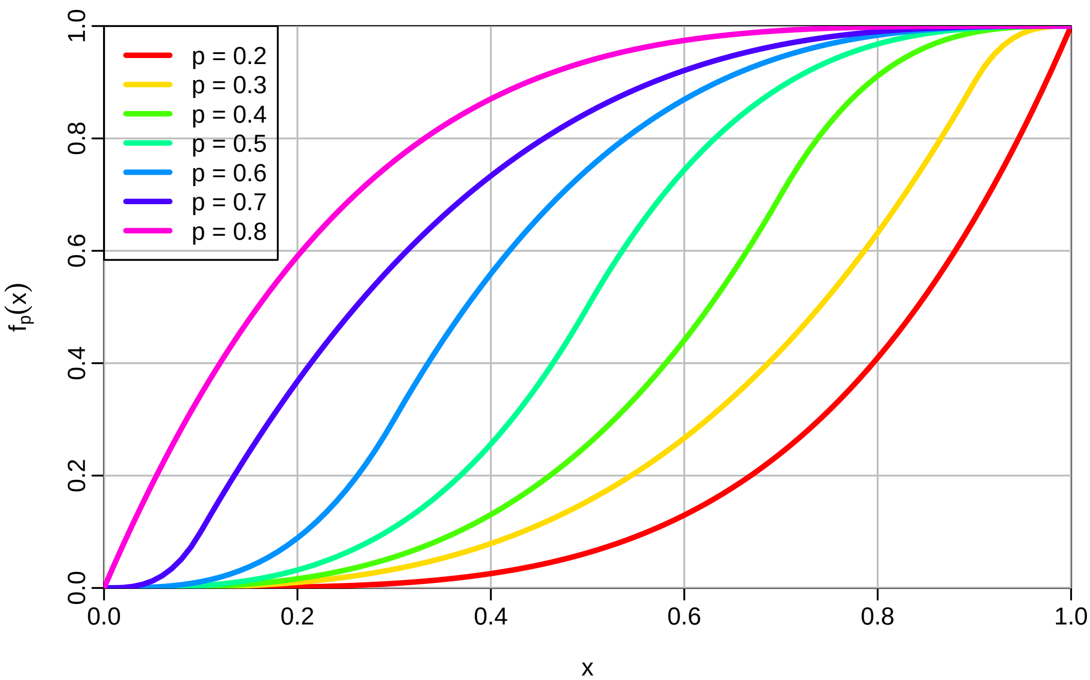


**Figure 1:** *Participation curves for p between 0.3 and 0.8.*

This family of functions has the following properties:

- The curves go through the points (0,0) and (1,1).
- The curves are continuous and monotonically increasing in both $x$ and $p$.
- The area under the curves is equal to $p$ in all cases.
- For $0.25 \leq p \leq 0.75$, the curves are cubic spline functions with a single knot at $(\alpha, \alpha)$.
- The curves are sigmoidal in shape for $0.25 \leq p \leq 0.75$.

# 3 Matching 2014 data

In this section we compare, for each Tertiary Admission Centre (TAC), the ITI values that actually occurred in 2014 (ACTAC, 2015), with the theoretical values from this participation model (i.e. the values that would have occurred if the above cubic spline model had been used).

Students are allocated an ITI within bands of width 0.05, from 0.00 to 99.95. Thus, there are 2000 ITI bands. To reduce the variation in the data, the 2000 ITI bands were aggregated into 200 ITI bands, each of width 0.50. The resulting numbers of students in these larger ITI bands are shown in Figure 2.

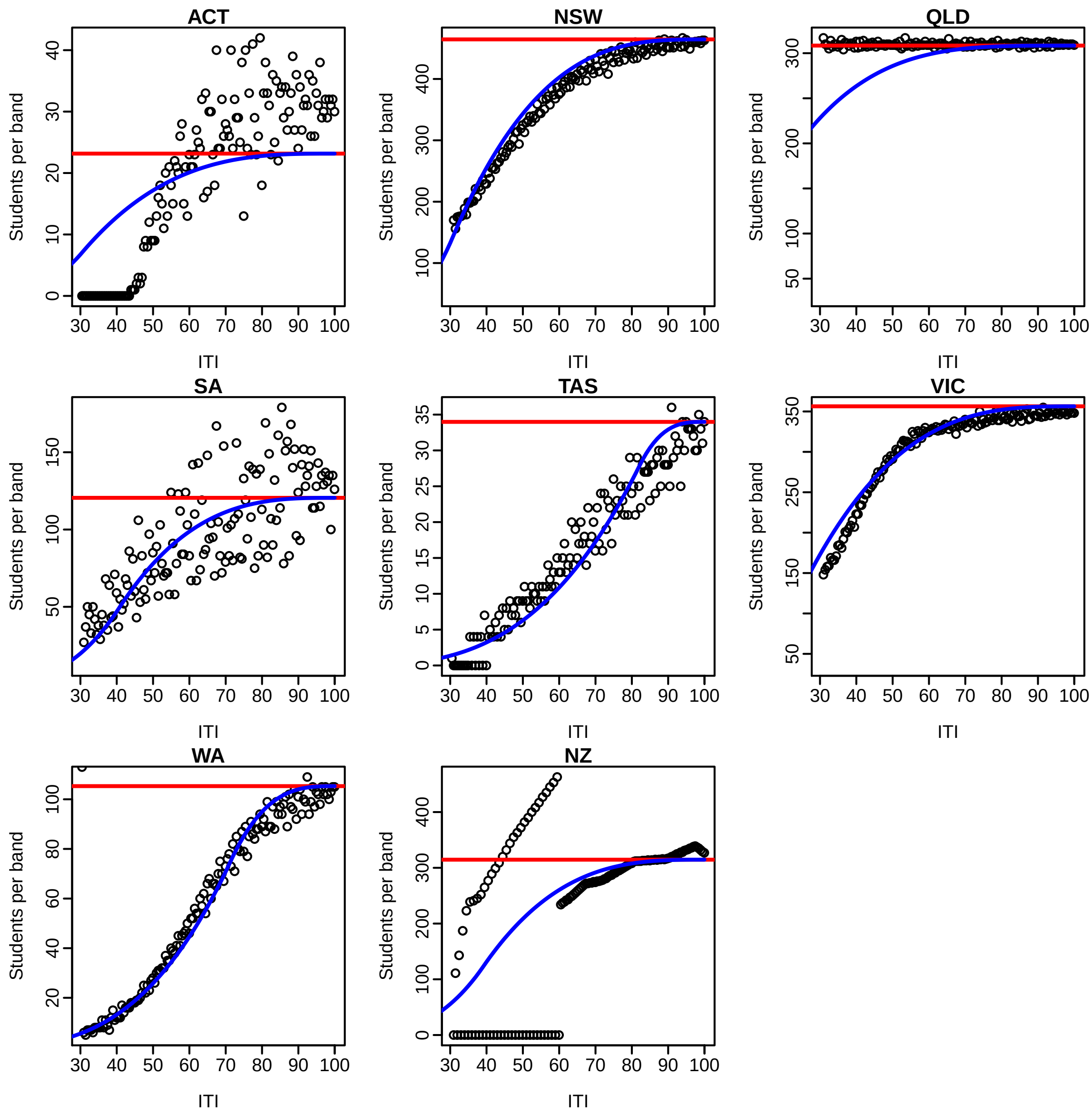


**Figure 2**

If there was a 100% participation rate, then we would expect 0.5% of the weighted cohort population to fall within each ITI band. The red line in each panel of Figure 2 shows this value for each state. The blue line shows the number of students that would have been in each ITI band if our proposed model had been followed.

The problems with ACT, Queensland and New Zealand were previously documented in Hyndman (2015) and will not be discussed further here. For the remaining states, there is close agreement with our proposed model.

Another way to look at the data is to consider cumulative percentages; these are the percentages of ITI eligible cohorts that receive ITIs in or above a given ITI band. We are particularly interested in differences between the actual and the theoretical cumulative frequencies.

Figure 3 shows that actual relative cumulative frequencies agree quite closely with theoretical ones for most TACs. This is particularly true for NSW and VIC, where the deviations are less than 4% over the entire ITI range and less than 2% for ITI $>$ 60%. SA and NZ slightly exceed the theoretical values at the higher ITI levels and are lower at lower ITI levels. The reverse is true for TAS, and for WA the actual values are slightly less than the theoretical ones across the entire range. The actual values in QLD exceed the theoretical ones across the entire ITI range, and the excess increases as ITI decreases.

Figure 4 show the differences between actual and theoretical relative cumulative frequencies at the critical top end of the ITI range.

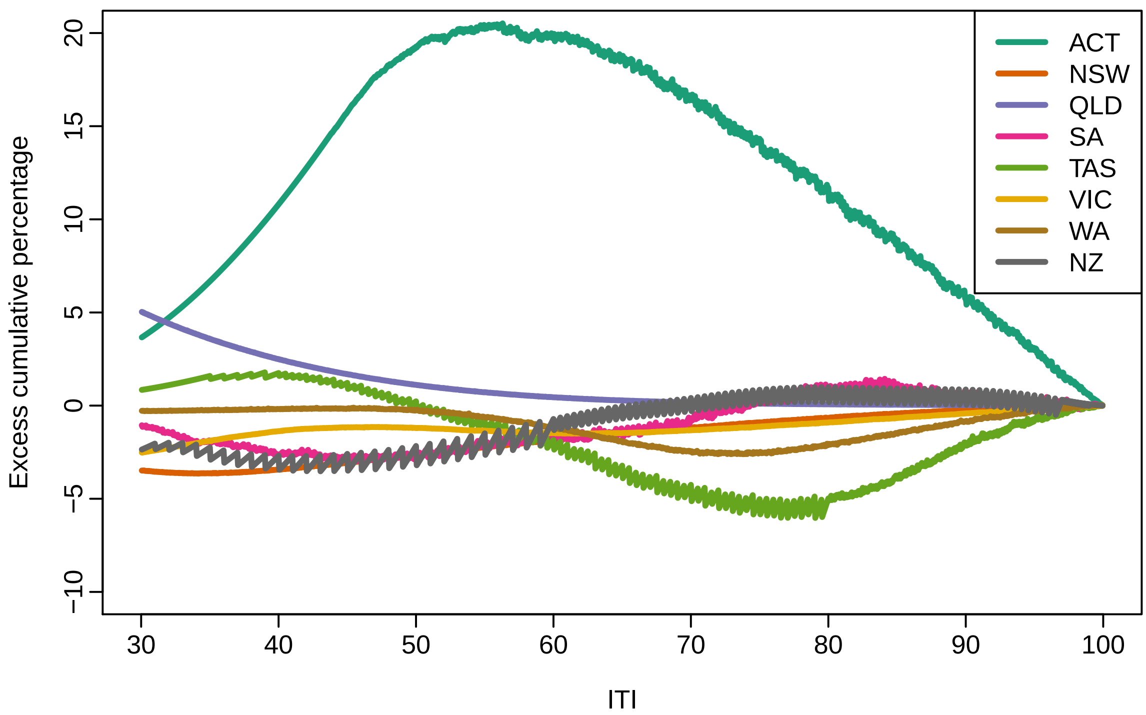


**Figure 3:** *Excess cumulative percentages per state under the cubic spline model.*

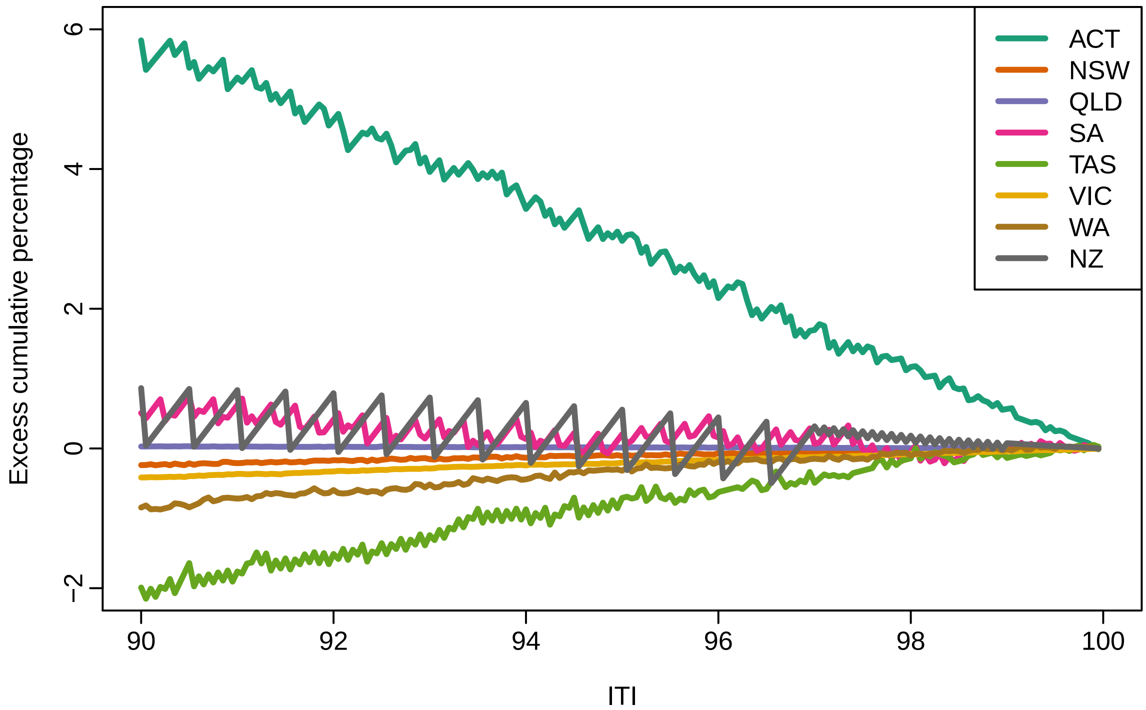


**Figure 4:** *Excess cumulative percentages per state for ITI > 90 under the cubic spline model.*

## 4 Summary

We have shown that a one-parameter family of cubic splines provides models of participation curves that give distributions of ITIs that closely match the actual distributions for most TACs. The reason for the close match is due to the sigmoid nature of the cubic spline graphs, a property shared with logistic functions and the few 'actual' participation curves that are available.

The single parameter is determined uniquely by the overall participation rate, and there is reasonably good agreement about how this is defined. Since the computations involved are straightforward, the use of cubic splines to model participation curves is both simple and transparent.

Brown (2015) makes a strong case for the need to validate any choice of method of producing participation curves with appropriate data, and we concur. It should be noted that the graphs of the cubic splines we propose are very similar in shape to the graphs of logistic functions, and they are entirely based on known data (the participation rates).

We believe that it is most unlikely that the rankings based on Year 9 NAPLAN results are so closely correlated with Year 12 ITI rankings that their use will lead to significantly better theoretical participation curves. Of course, this assertion can, and should, be tested. In the meantime, we suggest that it better to use the simpler model that we propose here.

## Appendix: Mathematical discussion

Our participation model is given by (1). For $0.25 \leq p \leq 0.75$, this consists of a piecewise cubic function with the pieces joining at $(\alpha, \alpha)$. The function, and its first two derivatives, are continuous at $(\alpha, \alpha)$. Thus, it is a cubic spline with a single 'knot' at the point $(\alpha, \alpha)$ which lies on the line $y = x$.

If $f_p(x)$ is the participation function, then the overall participation rate is the integral $\int_0^1 f_p(x)dx$. Elementary calculus shows that, for $0.25 \leq p \leq 0.75$,

$$\int_0^1 f_p(x)dx = \int_0^\alpha \frac{x^3}{\alpha^2}dx + \int_\alpha^1 \left(1 - \frac{(1-x)^3}{(1-\alpha)^2}\right)dx = \frac{\alpha^2}{4} + (1-\alpha) - \frac{(1-\alpha)^2}{4} = \frac{3}{4} - \frac{\alpha}{2} = p.$$

Similarly, for $p < 0.25$ and $p > 0.75$, it is easy to show that $\int_0^1 f_p(x)dx = p$.

Although the function definition changes at $p = 0.25$ and $p = 0.75$, there is a smooth transition at these points. Note from (1) that when $p = 0.25$, $\alpha = 1$ and $f_p(x) = x^3$. Similarly, when $p = 0.75$, $\alpha = 0$ and $f_p(x) = 1 - (1-x)^3$.

Some TACs choose the second parameter for the logistic model so that $x_M$, the ITI corresponding to a 50% participation rate ($f_p(x_M) = 0.5$), is a predetermined value. There is no such freedom with the cubic spline model. Elementary algebra shows that with a cubic spline, $x_M$ and $p$ are related as follows:

$$x_M = \begin{cases} 2^{p/(p-1)} & \text{if } 0 \leq p \leq 0.25, \\ (3-4p)^{2/3}/2 & \text{if } 0.25 \leq p \leq 0.5, \\ 1-(4p-1)^{2/3}/2 & \text{if } 0.5 \leq p \leq 0.75, \\ 1-2^{1-1/p} & \text{if } 0.75 \leq p \leq 1, \end{cases} \tag{4}$$

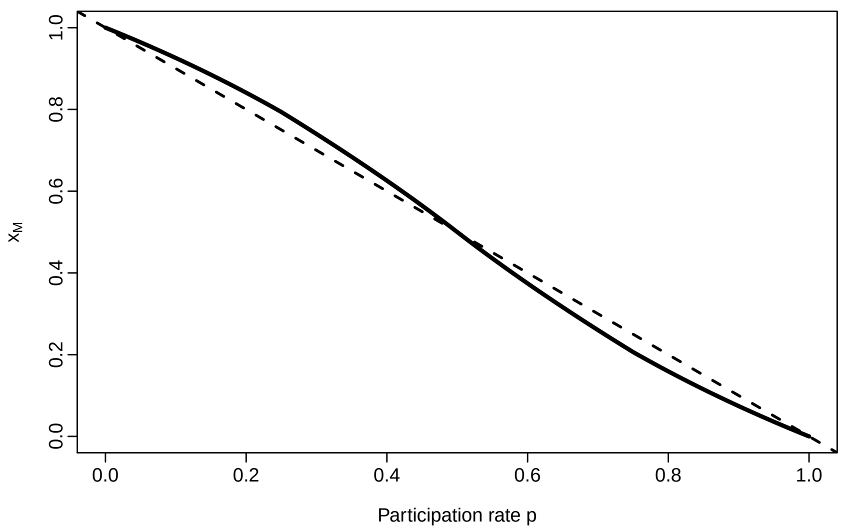


**Figure 5:** *The solid line shows $x_M$ as a function of the participation rate $p$. The dashed line is $1 - p$.*

These are rather complicated and uninformative expressions. However the relationship between $x_M$ and $p$ is approximately linear. In fact $x_M$ differs from $1 - p$ by less than 5% over the entire range of $p$ values.

## References


ACTAC (2015). Consolidated ITI distributions 2014.

Brown, T. C. (2015). Review of the Interstate Transfer Index methodology.

Harrison, K. J. (2014). The use of power curves for calculating the ITI.

Hyndman, R. J. (2015). Analysis of 2014 ITI data.

NSW Universities Admission Centre (2014). ITI calculation report, NSW 2014.