Relating ACT and NSW UAI populations via PISA and other scores

1. Introduction and Overview

Up until 1976, ACT school students seeking tertiary admission took the NSW HSC examination, and so gained tertiary admission qualifications on the same academic footing as their NSW contemporaries. In terms of academic performances (i.e. educational measurements), *the distribution of HSC aggregate scores of the ACT students was effectively the same as the distribution of their NSW contemporaries* [these observations are due to Morgan for 1975 data, and Daley's analysis of 1976 HSC results supplied by Mitchell]. In other words, the selection mechanism by which students sought a tertiary admission credential in NSW and the ACT, produced across NSW on the one hand and within the ACT on the other, two candidatures of approximately the same spread of academic ability, notwithstanding the higher proportion of the age cohort (about 45%) in the ACT compared with about 30% in NSW.

Since 1976, the ACT and NSW systems have diverged somewhat: the curriculum choices available to ACT students fairly quickly became wider in their offerings, and the students responded by choosing more broadly, outside the strictures of a predominantly academic preparation. While the proportion of the age cohort completing a university admission credential has grown in the ACT (from about 45% to about 55%), there has been a larger growth (up to about 90% of the age cohort) receiving a Year 12 Certificate (this is often loosely interpreted as 'completing Year 12'). The corresponding population figures for NSW are somewhat different: there, the majority (loosely, 85 to 90%) of their HSC students (about 65 to 70% of the cohort) seek UAI. This constitutes substantial growth from around 30 to 35% of the age cohort obtaining HSC in 1975–76.

It is possible in principle to check the equivalence of NSW HSC-based UAIs and ACT Tertiary Entrance score-based UAIs via first year university performance, provided there are sufficient numbers of students with both credentials taking common courses, but such predictive validity studies are plagued by other factors affecting their interpretation.

It is more feasible (and in principle simpler) to use the larger school-based populations whenever the students have been assessed by some common instrument, subject to having adequate information on the subsequent pathways of the populations. Since 1976, to DJD's knowledge there has not been any use of a common test of both ACT and NSW students yielding educational measurements that could be used to establish relative academic achievement of ACT and NSW assessments at the Year 12 level, until there recently became available the stratified sample-based data of the PISA (Programme for International Student Assessment; see *PISA in Brief from Australia's Perspective*, `www.acer.edu.au`). These data, of which a more comprehensive account is given in the full Australian report *Facing the Future* compiled after testing 15 y.o.s in 2003, demonstrate clearly through their analyses by the several Australian States and Territories, that the ACT 15 y.o. population outperforms almost all other states (and in particular, New South Wales) in all of Mathematical, Reading and Scientific Literacy, and Problem Solving. This means that the computation of Interstate comparability based on an assumption that Year 10 populations ($\approx$ 15 y.o. population) in the several systems have similar distributions of academic ability, is clearly false for the ACT, as long noted by the ACT system and recognized in the way that ACT data are merged with NSW data for use by the NSW and ACT Universities Admission Centre. (Adams (1984) *Sex Bias in ASAT?*, ACER, also has summaries of three years' data concerning 'UAI' populations in ACT, WA and Queensland, again showing that the ACT population differs from those in other states.)

There are also available within the ACT the results of Literacy and Numeracy (LitNum) test results of Year 9 students together with identification of those students at Year 12 who have taken the test in Year 9. Within NSW there are available Year 10 School Certificate results together with identification of those students two years later at Year 12 (and taking the NSW HSC examination).

This assumption of the equivalence of Year 10 populations in different parts of Australia in terms of their academic performance, was the basis of an agreement between the states c.1995 to provide comparability of Tertiary entrance ranks for immediate school leavers across the different state systems, despite the assumption being mostly untested and largely unquestioned. It is arguable that detailed academic comparability is not necessary within a system where education, when regarded as a community asset and service, is a responsibility of the region (state) it is serving and not the Commonwealth. Nevertheless, it would appear to be a self-evident principle that within Australia, a student's tertiary entry credential should be acceptable on the same terms both within and beyond the region where it is issued.

2. Data and Assumptions

Our first task is to record idealized assumptions on which the PISA data might be used to relate the academic performances of the ACT and NSW UAI populations when coupled with ACT LitNum data on the one hand, and NSW SC/HSC data on the other. Note that observed scores are treated as true scores, in order to establish routines that might be valid in an ideal world. In practice, observed scores lead to appreciable biases as previously indicated to the NSW Technical Committee on Scaling. I have not yet had time to examine the (anticipated) second-order but systematic effects in relation to the results below of computations based on observed scores without allowance for such effects. What is relevant to note here is that such biases vary between populations with different score distributions.

I make the following specific assumptions concerning the three sets of data (or, of statistics based on those data):

(A.1) The ACT LitNum data, consisting of a frequency table of aggregated Literacy and Numeracy scores of a Year 9 population in 2002, subsequently thinned to those who in December 2005 received either an actual or a notional UAI, gives an estimate of

$$q_{\text{ACT}}^{\text{UAI}}(x) = \Pr\{\text{ACT Yr.9 rank } x \ (0 < x < 1) \text{ receives UAI}\},$$
$$q_{\text{ACT}}^{\text{Y12}}(x) = \Pr\{\text{ACT Yr.9 rank } x \ (0 < x < 1) \text{ receives actual or notional UAI}\}$$

(strictly, 'ACT Yr.9 rank x' is shorthand for 'an ACT student with ACT Yr.9 rank x based on the LitNum test data', and 'rank x' is a shorthand for 'true rank x' in the Year 9 population [this assumes that such a ranking exists]). The proportions of the three groups in terms of the LitNum scores are shown in Figure 2a: ACT Yr.9 as defined (——), the subpopulation with some actual or

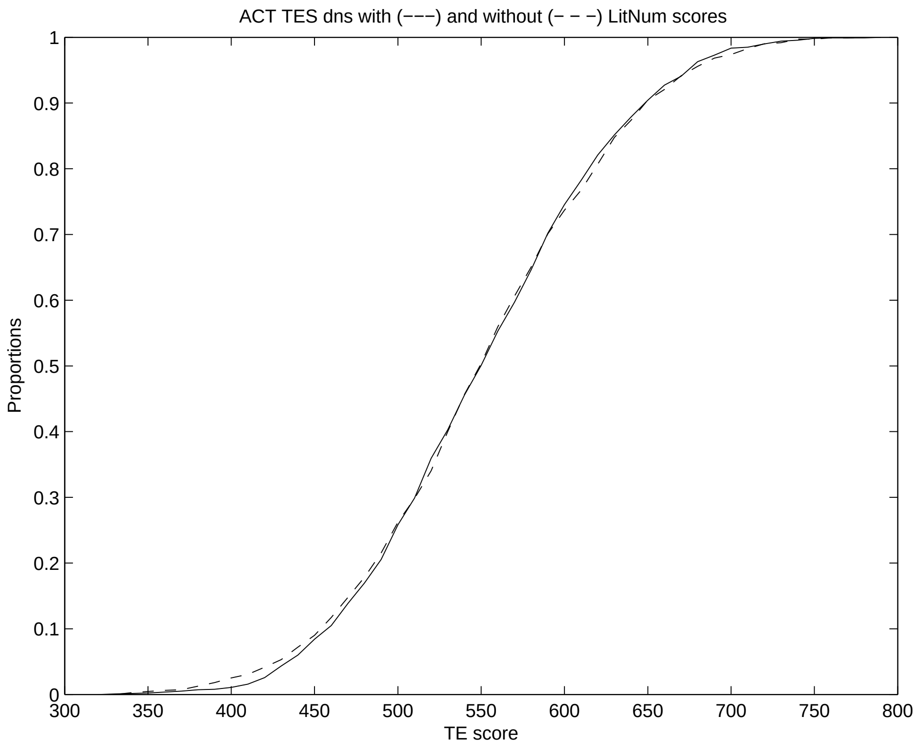

Figure 1: ACT 2005 TES dns with (——) / without (– – –) Year 9 2002 LitNum scores.

notional UAI score at the end of 2005 (— · — · —), and the further subpopulation receiving an actual UAI (– – –). Note that the ACT Year 12 population of over 4000 in 2005 included at least 300 students from NSW, over 100 international students, and a significant number (between 100 and 200, say) of other new arrivals to the ACT: the actual number of students with LitNum scores who later obtained a UAI comprised about 60% of those obtaining a UAI through the ACT system. The distribution of ACT TE scores in 2005 of the two groups (viz. with and without LitNum scores) is shown in Figure 1: for practical purposes the distributions are the same (but, see also Figure 8 in Section 4). [In the Year 10 School Certificate candidature in NSW, students who are not in the NSW HSC candidature two years later comprise at least 2 or 3% of the candidature at any level of SC-attainment.]

The data underlying Figure 2a, comprise all Year 9 students with at least one component of the Year 9 LitNum test, plus selected substitute school populations as estimates for the three schools not participating in that test in 2002. Thus, the data are regarded as a good or better estimate of the total ACT population, at least as reliable as the PISA-based estimates of the ACT population.

(A.2) The *Report on Scaling 2005 NSW HSC* (short title for the UAC publication), at its Figure 3.2 (based on its Figure 3.1), plots proportions that are interpreted probabilistically as

$$p_{\mathrm{NSW}}^{\mathrm{UAI}}(z) = \Pr\{\text{NSW Yr.10 score } z \text{ receives UAI}\}.$$

Equivalently, these data can be converted to quantile form so that we then have available

$$q_{\mathrm{NSW}}^{\mathrm{UAI}}(x) = \Pr\{\text{NSW Yr.10 rank } x \text{ receives UAI}\}.$$

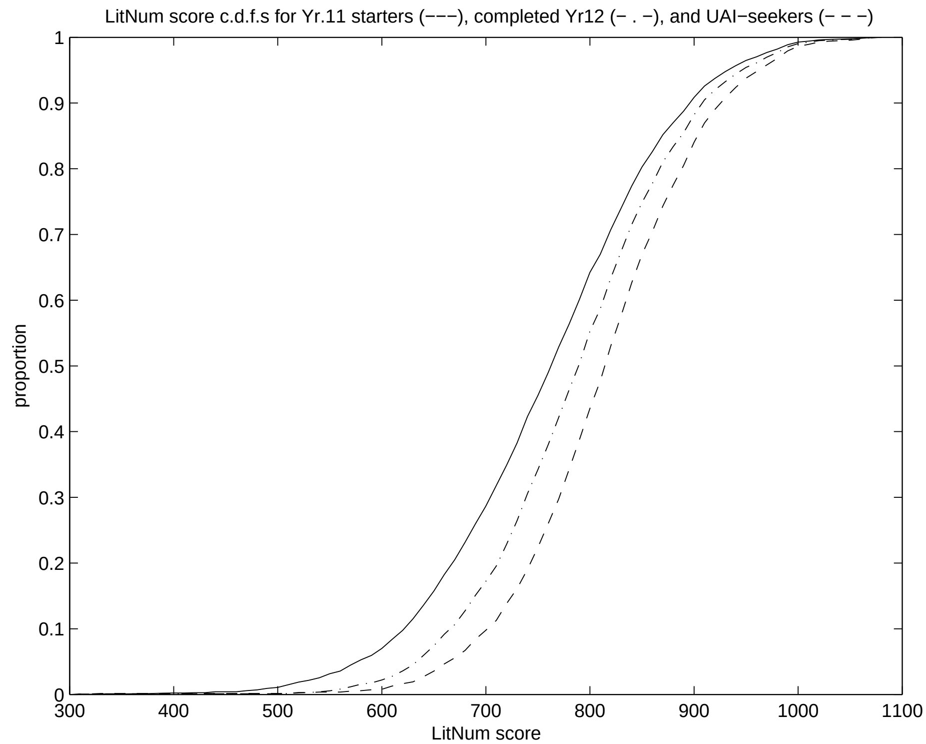

Figure 2a: ACT Year 9 (2002) LitNum data and their 2005 Year 12 participation.

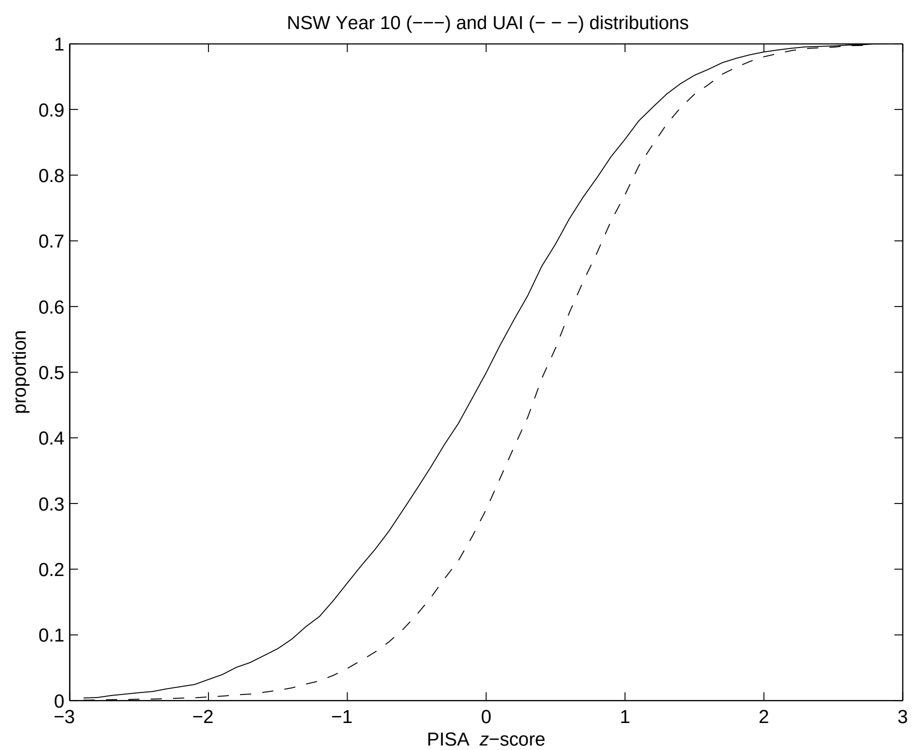

Figure 2b: NSW 2003 Year 10 (—) and 2005 UAI (− − −) distributions.

(A.3) The PISA-test based data give composite z-score results (constructed by ACER from the test results as an equally weighted combination of Mathematical and Reading sub-scales) for samples

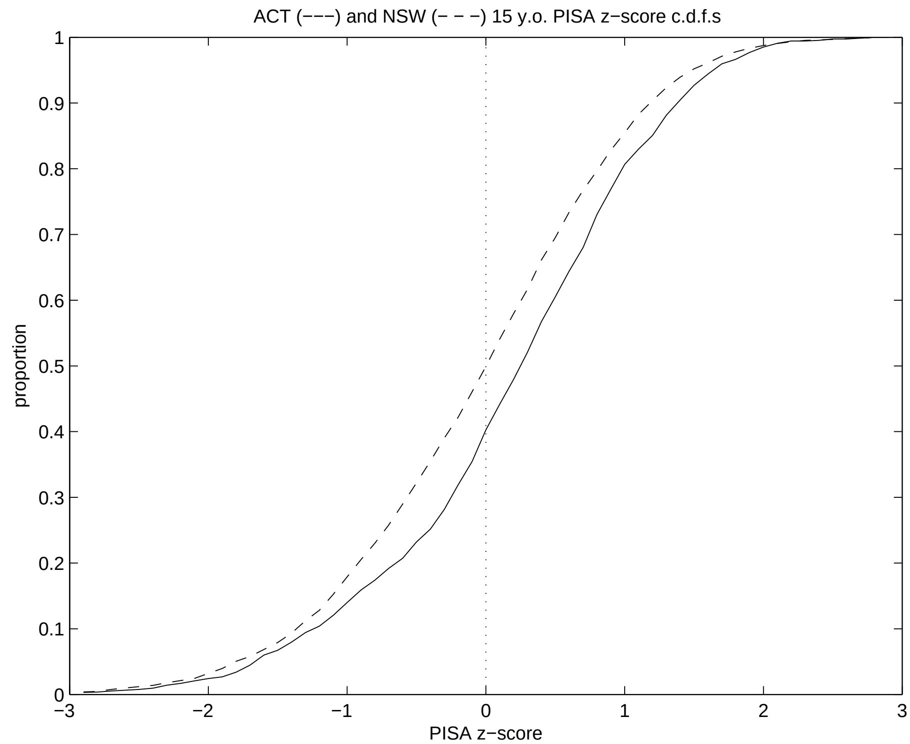

Figure 3: ACT (——) and NSW (– · – · –) Year 10 pop'n c.d.f.s of PISA z-score.

of students within schools and samples of schools by school-type (Government, Catholic, Other). When combined with weights (given with the data from ACER) that reflect this stratification, we can construct estimated frequency distributions of the *c*.15 y.o. populations of ACT and NSW on a common test. Assume that these populations are equivalent to the Year 9 or 10 populations in the two regions, so that the ACT academic performances in the aggregated LitNum tests in 2002 and on the PISA test in 2003 are equivalent, and in the aggregated NSW SC test and on the PISA test in 2003 are also equivalent. Denote these (cumulative) frequency distributions by F_{ACT} and F_{NSW} respectively, so that e.g.

$$F_{\mathrm{ACT}}(z) \;=\; \text{proportion of ACT students with PISA-score} \leq z.$$

These data are plotted in Figure 3, (——) for F_{ACT}, (– – –) for F_{NSW}. It shows for example that about 40% ACT students and 49% NSW students scored 0 or less (on the z-scale). These plot-points are consistent with graphical data in *PISA in Brief from Australian Perspective*.

Remark 2.1. It cannot be emphasized too strongly that the order of any differences between NSW and ACT can be expected to be affected seriously by the use of observational data without correction or parameterization (or other smoothing). The formulae used above are valid (under the assumptions as indicated) for distributions of *true scores* rather than of observed data, and the noisy nature of each individual measurement does not "vanish in the large sample limit" (i.e. by using datasets from larger populations) because every measurement has large and significant 'error', and the functions $q(z)$ vary, monotonically, with z (this monotonic variation in participation then leads to a bias effect). On the other hand, the fact that the observational data are based on

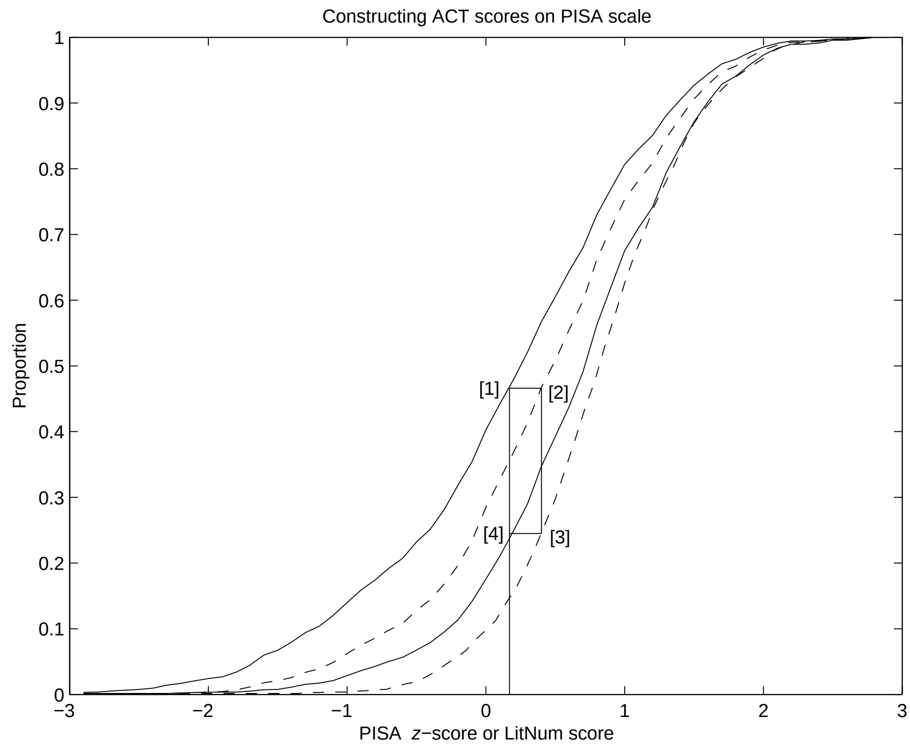

Figure 4

Illustrating construction of ACT UAI dist'n on PISA scale (——, lower) from 15 y.o. PISA dist'n (——, upper) via LitNum Yr.9 (– – –, upper) and UAI (– – –, lower) dist'ns (see Section 3).

responses to the same measurement instrument (i.e. same test), means that some of these effects should be 'common', though because the proportional retention functions $q^{\mathrm{UAI}}(\cdot)$ vary between ACT and NSW implies that different bias effects are introduced.

3. Computations

We are interested in using the data available to compare

$$G_{\mathrm{ACT}}(z) \;=\; \text{proportion of ACT UAI-seeking students with PISA scores} \le z,$$

with its NSW analogue $G_{\mathrm{NSW}}(z)$. To do this, we use two curves derived from the ACT LitNum data described in (A.1), namely

$$\begin{aligned}
F^{\mathrm{LN}}_{\mathrm{ACT}}(x) \;&=\; \text{proportion of all ACT Year 9 students with LitNum scores} \le x,\\
G^{\mathrm{LN}}_{\mathrm{ACT}}(x) \;&=\; \text{proportion of ACT UAI-seeking students with LitNum scores} \le x,\\
G^{\mathrm{LNY12}}_{\mathrm{ACT}}(x) \;&=\; \text{prop'n of ACT students with actual or notional UAI and LitNum score} \le x.
\end{aligned}$$

(To obtain a UAI, a student must satisfy minimum T-level course requirements; scores of students attempting some T-level course but less than the minimal package are used administratively to construct 'notional' TE scores and thereby produce the larger sub-population [cf. Figure 2a]).

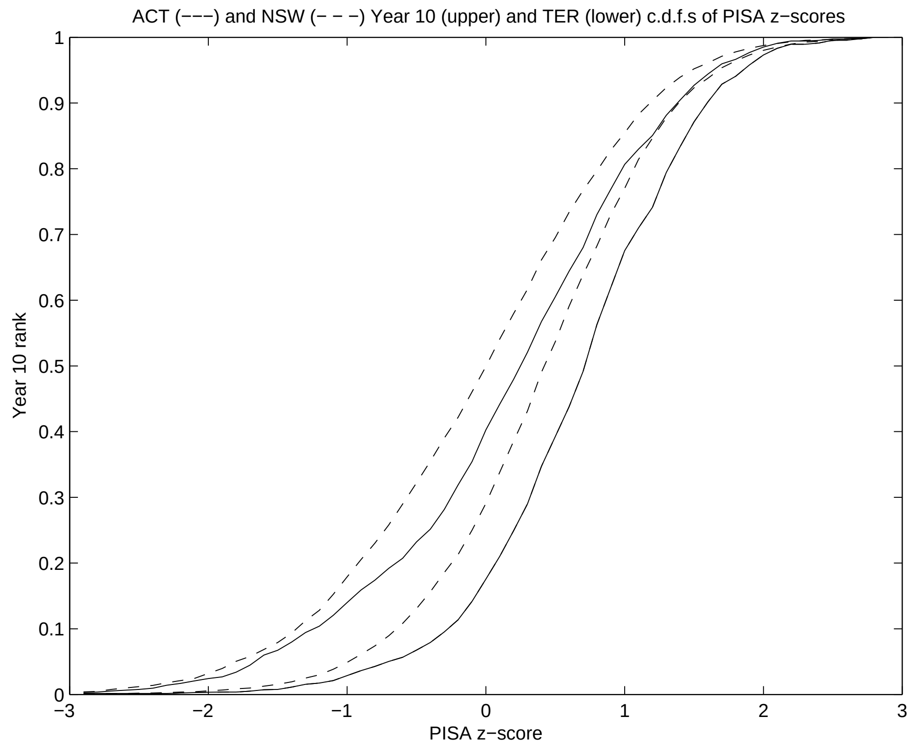

Figure 5

Proportions of ACT (——) and NSW (– – –) students with PISA z-scores:
(a) all *c*.15 y.o. (upper), and (b) UAI-seeking (lower).

We make the assumptions

(I) that the populations used to generate the ACT LitNum data in (A.1) and the ACT PISA data in (A.3) are similar in the sense that, if only their measurements were on the same scale, the distributions would be the same, so that, if only the distributions were of true scores, there would be a one-to-one mapping between PISA z-scores and LitNum aggregate scores, and therefore the group of students with PISA scores $\leq z$ would have the same 'achievements' as the group with LitNum scores $\leq$ some x_z, say: this x_z would be obtainable as the solution of $F_{\text{ACT}}^{\text{LN}}(x_z) = F_{\text{ACT}}(z)$, and so can be written (in inverse function notation)

$$x_z = \left(F_{\text{ACT}}^{\text{LN}}\right)^{-1}(F_{\text{ACT}}(z));$$

and

(II) that the observed score distributions can be regarded as the true score distributions (i.e. there is no noise in either distribution).

Granted these assumptions, and using the notation as given, a PISA-score z is equivalent to the ACT LitNum score x_z, and therefore, since $G_{\text{ACT}}^{\text{LN}}(x)$ is the proportion of ACT UAI-seeking students with LitNum scores $\leq x$, the desired proportion of ACT students with PISA scores $\leq z$ is just

$$G_{\text{ACT}}(z) \;=\; G_{\text{ACT}}^{\text{LN}}(x_z) \;=\; G_{\text{ACT}}^{\text{LN}}\Big(\left(F_{\text{ACT}}^{\text{LN}}\right)^{-1}(F_{\text{ACT}}(z))\Big).$$

The lower continuous curve in Figure 5 is obtained by treating the observed score distributions as these idealized true score distributions and thus measurement errors in the various test data have been ignored.

How this mapping is used numerically in practice is indicated by the box in Figure 4. Start by constructing a vertical line through $(z, 0)$ representing possible proportions for a given PISA score z, and find the intersection [1] of this line with the upper continuous curve (= the 15 y.o. PISA distribution)—this intercept is the point $(z, F_{\text{ACT}}(z))$. On the horizontal line through [1] find the intersection [2] with the upper dashed curve comprising $F^{\text{LN}}_{\text{ACT}}(x)$ for LitNum scores x on a convenient scale, so [2] is the point $(x_z, F_{\text{ACT}}(z)) = (x_z, F^{\text{LN}}_{\text{ACT}}(x_z))$. On the vertical line through [2] find the intersection [3] with the lower dashed curve of the LitNum distribution $G^{\text{LN}}_{\text{ACT}}(\cdot)$, so [3] is the point $(x_z, G^{\text{LN}}_{\text{ACT}}(x_z))$. Finally on the horizontal line through [3] find the intersection [4] with the original vertical line, so that [4] is the point $(z, G^{\text{LN}}(x_z))$. The collection of all such points $(z, G^{\text{LN}}(x_z))$ is shown as the lower continuous curve in Figure 4, and repeated as the lower continuous curve in Figure 5.

A similar transformation using NSW data in conjunction with (A.2) yields proportions of NSW UAI students with PISA-scores $\leq z$. It is in fact available from the conversion mapping between TER and UAI percentiles supplied to ACT BSSS in late November 2005. The result of this mapping is the lower dashed curve shown in Figure 5.

From inspection of the PISA z-scale distributions of these two UAI-seeking populations, interpreted as subgroups of their respective Year 9 or Year 10 populations (called c.15 y.o. populations in the *PISA* document), it is evident that **neither of the pairs of ACT and NSW populations, namely the Year 9 and 10 populations, and the UAI-seeking Year 12 populations, is an academically equivalent pair.**

On the other hand, in 1975 and 1976, the ACT and NSW HSC examination populations were academically equivalent, and the assumption has been made since that time that the tertiary-entry qualified candidature at the NSW HSC examination, is academically equivalent to some ACT sub-population that is at least as large as the ACT Tertiary Entry Score qualified population; in 1985 this population was defined to consist of students who were either TES-qualified or had taken at least one T-level course and completed Year 12, because it had been realized that the proportion of the NSW age cohort taking the HSC had grown much more markedly than for the ACT TES-qualified cohort. The action taken then was not argued on educational measurement criteria as we are now doing.

> *The pertinent matter is to construct an ACT population that is approximately academically equivalent to the NSW UAI-seeking population.*

The fact that in terms of PISA z-scores the distribution of the NSW UAI-seeking population is generally below the distribution for the ACT Year 10 but above the distribution for the ACT UAI-seeking population (see Figure 5), implies that some intermediate ACT sub-population can yield a distribution that coincides with the NSW distribution.

Such a modified ACT sub-population would be academically equivalent to the NSW UAI-seeking population so that a given point on the z-scale would correspond to the same UAI percentile rank in the NSW population. It requires us to adjoin to the ACT population sufficient students from a group with on average lower z-scores.

The current ACT procedure is to adjoin those students who complete Year 12 with at least one T-subject but who do not have a complete T-package. This is a subgroup of the students who

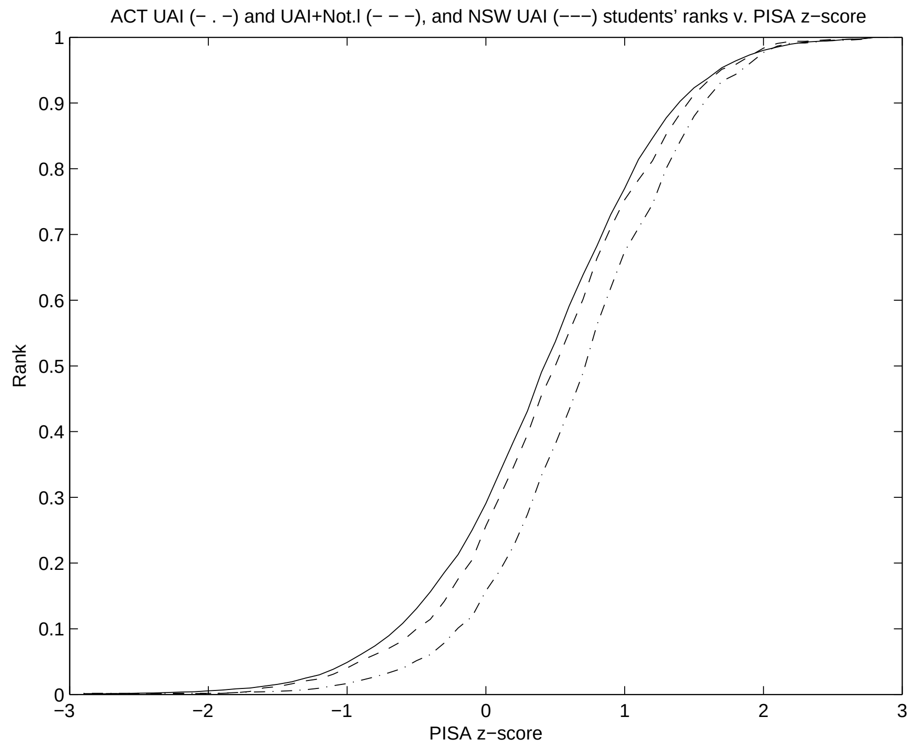

Figure 6

ACT UAI (− · −), modified UAI (− − −), and NSW UAI (——) pop'n c.d.f.s of PISA z-scores.

complete Year 12, and those in this last-named subgroup are identified in the LitNum data, so we can at least compare the effect on Figure 5 of amending the population whose c.d.f. of z-scores is plotted as the lower curve (——). The amended c.d.f., namely of

$$G_{\text{ACT}}^{\text{LNY12}}\Big(\Big(F_{\text{ACT}}^{\text{LN}}\Big)^{-1}(F_{\text{ACT}}(z))\Big),$$

is shown in Figure 6, together with the same NSW c.d.f. as in Figure 5. (The mechanics to construct this modified ACT curve are exactly as sketched in connection with Figure 4, except that $G_{\text{ACT}}^{\text{LNY12}}$ replaces $G_{\text{ACT}}^{\text{LN}}$.)

4. Discussion

Examination of Figure 6 indicates that the amended ACT population—if we accept for the moment that the various assumptions we have made en route hold for the data as used—is still marginally disadvantaged by being given a direct, population-based mapping of UAIs from the NSW base. The justification for this statement is that the ACT (actual + notional)-UAI c.d.f. in Figure 6 lies on or below (and mostly, below) the NSW c.d.f. over almost the entire range of the c.d.f. scale, albeit at a markedly smaller discrepancy than for the actual UAI-seeking population.

In other words, on the basis of the analyses presented above, the adjustment procedure of 2005 would appear to have been marginally insufficient.

In view of the above, an expanded definition of students to be given notional TE scores (hence, notional UAIs) has been framed with the effect of increasing the proportion of students to be

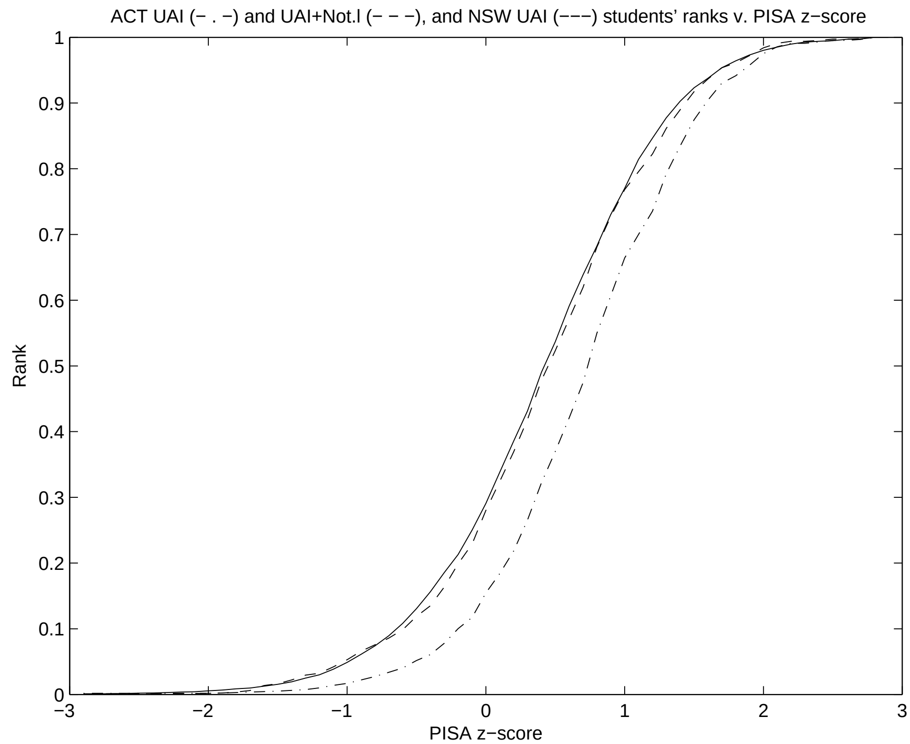

Figure 7

ACT UAI (− · −), expanded T-level (− − −), and NSW UAI (——) pop'n c.d.f.s of PISA z-scores.

adjoined to the UAI-seeking candidature. The results are shown in Figure 7; on the basis of the 2005 data, the expanded definition appears to be more satisfactory.

For 2006, the database of LitNum scores has only one sizable college for which there are no data. Recomputing the LitNum and No-LitNum score groups as in Figure 1 by reallocating the scores for colleges for which LitNum data are available in 2006, yields the pair of curves in Figure 8. Here, the match between the LitNum group (——) and No-LitNum group (− − −) is now no longer as close as in Figure 1. However, it is in the region of lower TE scores that the two groups now suggest a systematic deviation, and it is in precisely this area that the ACT and NSW curves in Figure 7 suggest a systematic deviation, in the reverse direction in the sense that, were the ACT population in Figure 7 enlarged by more students with on average lower PISA z-scores, the match of the ACT to NSW would be closer.

I conclude therefore that a population definition of an expanded definition of students to be given notional TE scores, applied to the whole ACT candidature (and not merely those with LitNum scores), would have given a better matching to NSW data for constructing UAI scores for ACT students on a par with NSW.

Acknowledgement. I thank Bob Edwards of ACT BSSS for discussion and assistance with provision of data. In particular, the PISA data came from the Australian Council for Educational Research where John Ainley and Wolfram Schulz were of particular assistance to the ACT BSSS.

Any opinions in this document are those of the author.

DJD, 6–7, 17–23.ix.06 and 6–20.xi.06

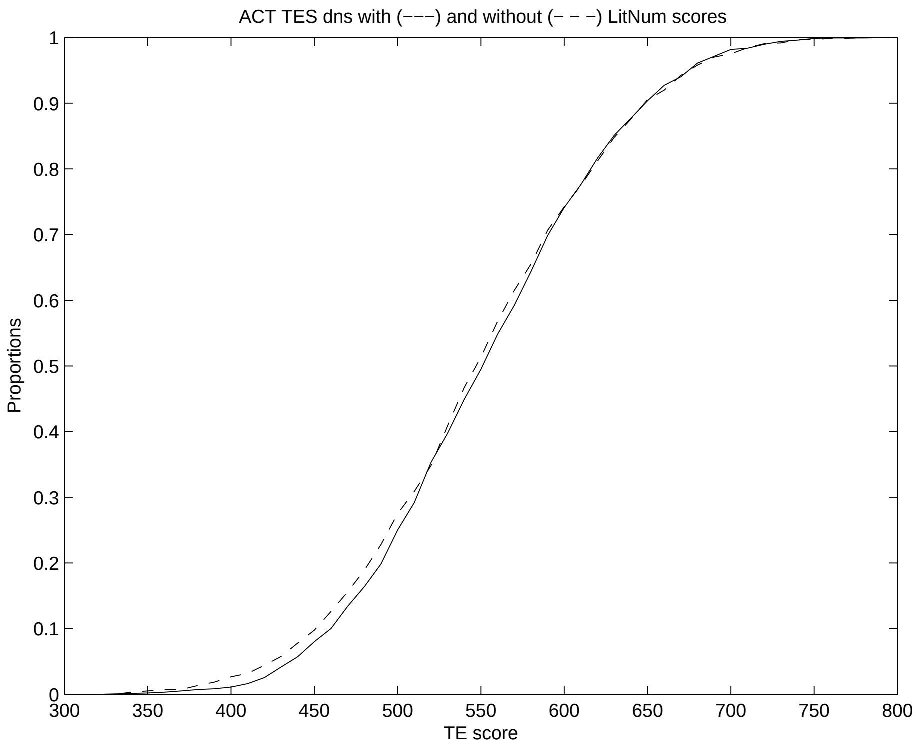

Figure 8. ACT 2005 TES dns, for 2006 LitNum (——) and No-LitNum (– – –) score groups.

APPENDIX: Miscellaneous Notes

The 2002 LitNum dataset as supplied in November '06 comprises data from 25 colleges/high schools, with three substantial schools (CGGS, MARC and MMKC, 709 pupils according to February 2002 School Census data) and five smaller schools (132 pupils) not participating. The dataset covers (as supplied) 1520 pupils who gained UAIs, 584 with notional TE scores, and 1569 with neither (but including 144 CBGS pupils with NSW HSC results, not relevant here). In terms of the School Census data, the discrepancies between LitNum participation and census numbers are in the range 2 to 10 per cent 'missing'. (For a Poisson distribution for missing students with rate 6%, schools of enrolments of 100 and 200 would yield 2SD limits (1,11) and (5,19), which roughly covers the observed discrepancies. I conclude that it is reasonable to regard non-participation rates as approximately random.)

Concerning missing schools, the 2006 dataset needs to be examined first to give some indication of the likely LitNum score performance of those that participated in 2003. This was done so that the distribution of LitNum scores in Figure 2a was obtained from the original dataset enlarged by using scores from schools with similar results to the 2006 LitNum scores for MARC and MMKC, and CBGS for CGGS. This procedure resembles, albeit more crudely (but on a much smaller scale), the computation of the 'full' ACT and NSW PISA c.d.f.s from the stratified sample data.