

***Growing Up in Australia: The Longitudinal Study of Australian Children***  
**Report on *Adapted PPVT-III* and *Who Am I?***

**Sheldon Rothman**

**Australian Council for Educational Research**

***Part A: Wave 1 scoring***

The first wave of the Longitudinal Study of Australian Children (LSAC) used two tests with the four-year-old sample. The *Adapted PPVT-III* is a shortened version of the *Peabody Picture Vocabulary Test, Third Edition* (Dunn & Dunn, 1997), which is a test of receptive vocabulary used as a screening test of verbal ability. This adaptation is based on work done in the United States for the Head Start Impact Study, with a number of changes for use in Australia. *Who Am I?* (de Lemos & Doig, 2000) assesses the cognitive processes that underlie the learning of early literacy and numeracy skills. One item was added to the standard *Who Am I?* booklet for use in LSAC. Summary statistics for each test are shown in Table 1.

**Table 1. Summary Statistics for Administration of Adapted *PPVT-III* and *Who Am I?* tests as part of LSAC Wave 1**

	<b>Adapted PPVT</b>	<b>Who Am I?</b>
Number of cases	4407	4827
Mean scaled scores	64.2 (se=0.123)	63.8 (se=0.125)
Mean number of items correct/mean raw score	28.2 (se=0.086)	25.6 (se=0.103)
Minimum number of items correct	2	0
Maximum number of items correct	40	44
Reliability	0.76	0.89

Note: For the *Adapted PPVT-III*, it was assumed that children who were not required to answer 10 ‘basal’ items had answered these items correctly. Reliability reported here is the person separation reliability (Wright & Masters, 1982).

***Adapted PPVT-III***

The PPVT-III was adapted for use in LSAC by altering the administration procedures, reducing the number of items administered during testing. To determine which items to retain for the adapted version, 215 children aged from 41 to 66 months (mean=54.7 months) were given the PPVT-III, with test administrators following standard procedures. After testing, a one-parameter (Rasch) item response model was fitted to the data, which consisted of correct and incorrect responses. The

person separation reliability was 0.88. After determining the ‘best’ 40 items for use in a shortened version, the remaining items were then fit again to a one-parameter item response model; the person separation reliability decreased to 0.78.

Development of the model suggested that 37 per cent of children would require only the core set of items, 5 per cent would require the core and basal sets, and 58 per cent would require the core and ceiling sets, resulting in an average of 26.3 items administered. The Pearson product-moment correlation between the full PPVT-III and the adapted PPVT-III was 0.93 for all children, and 0.91 for four year-olds.<sup>1</sup>

### *Scaling*

The adapted PPVT-III was scaled using a two-stage process. In the first stage, only the core set of 20 items was used, as these items had been administered to all children. For these core items, Rasch estimates were determined for each item, providing an indication of their difficulty. In the second stage, all 40 items were fitted, using the item estimates for the core items as anchors. This gave item estimates for the basal and ceiling items relative to the core items. The final case estimates were then transformed to a scale with a mean of 64 and standard deviation of 8.

### *Results*

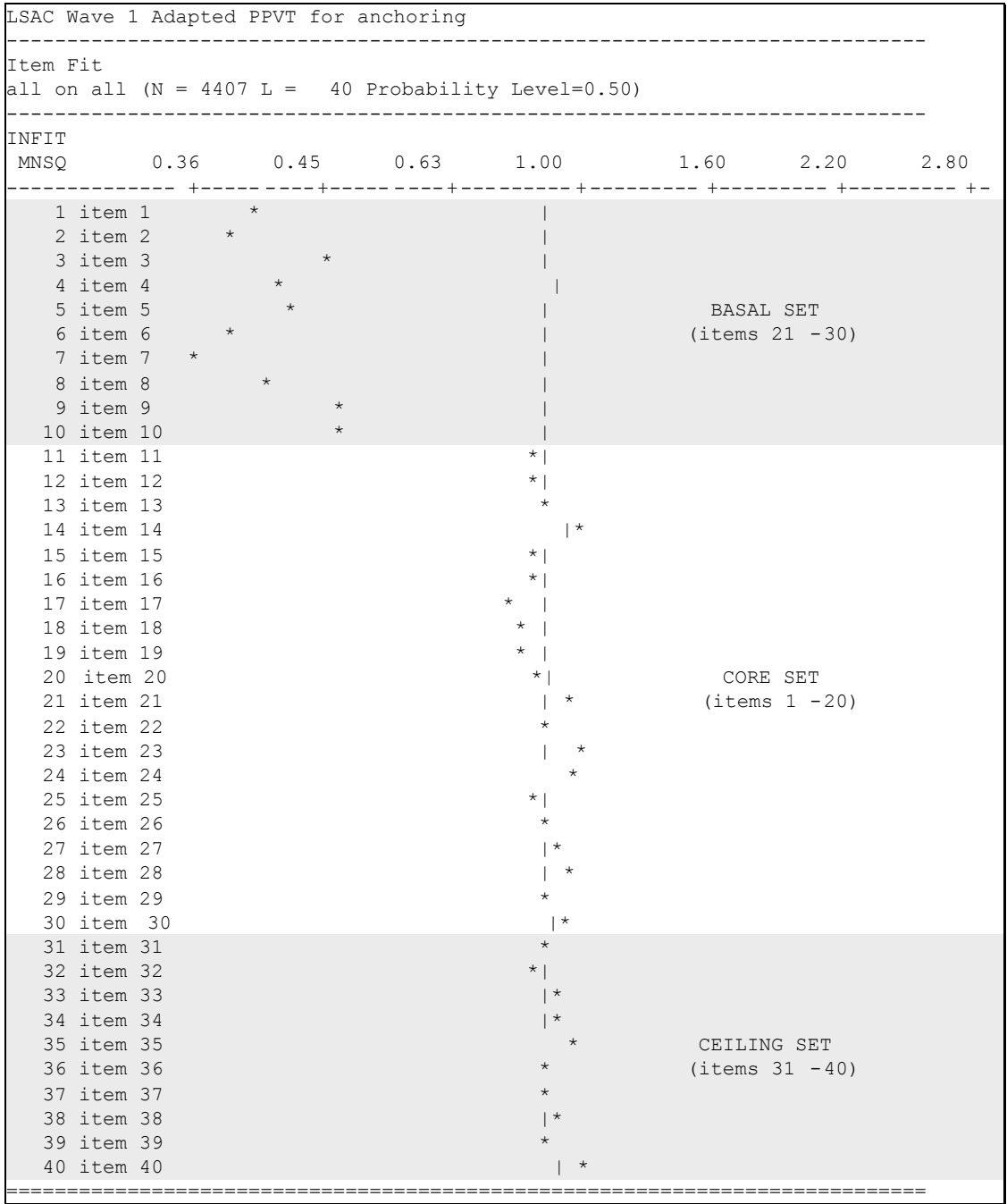
In Wave 1, 4407 children were administered the *Adapted PPVT-III*. Children ranged in age from 43 months to 79 months (mean=57.3); 21 per cent were aged 60 months or older. Twenty-one per cent of children were administered only the core set, 1 per cent the core and basal sets, and 78 per cent the core and ceiling sets, resulting in an average of 27.9 items. The test had a person separation reliability of 0.76.

### *Quality of the test*

The statistics indicate that the core and ceiling items used for the *Adapted PPVT-III* test fit the Rasch model well. This is shown in Figure 6, the item fit map. The infit mean square ranged from 0.86 to 1.17 for items 11–30 (the core set) and items 31–40 (the ceiling set). On each of the items in the basal set (items 1–10), the infit mean square was extremely low (0.49 or less) because only 30 children (1%) were administered these items; all other children were assumed to have correctly answered these items. The item map in Figure 6, which shows the item estimates (difficulties) mapped against the case estimates (children’s ability levels), shows that the basal items were appropriate for children given that set, but that the core and ceiling items were relatively easy for those who were given those sets.

---

<sup>1</sup> Rothman, S. (June 2003). *An Australian version of the Adaptive PPVT-III for use in research*. Unpublished paper. Melbourne: Australian Council for Educational Research.



**Figure 1. Item fit map for all items on the Australian adaptation of the Peabody Picture Vocabulary Test (PPVT-III) calibrated with all cases anchored to core items.**

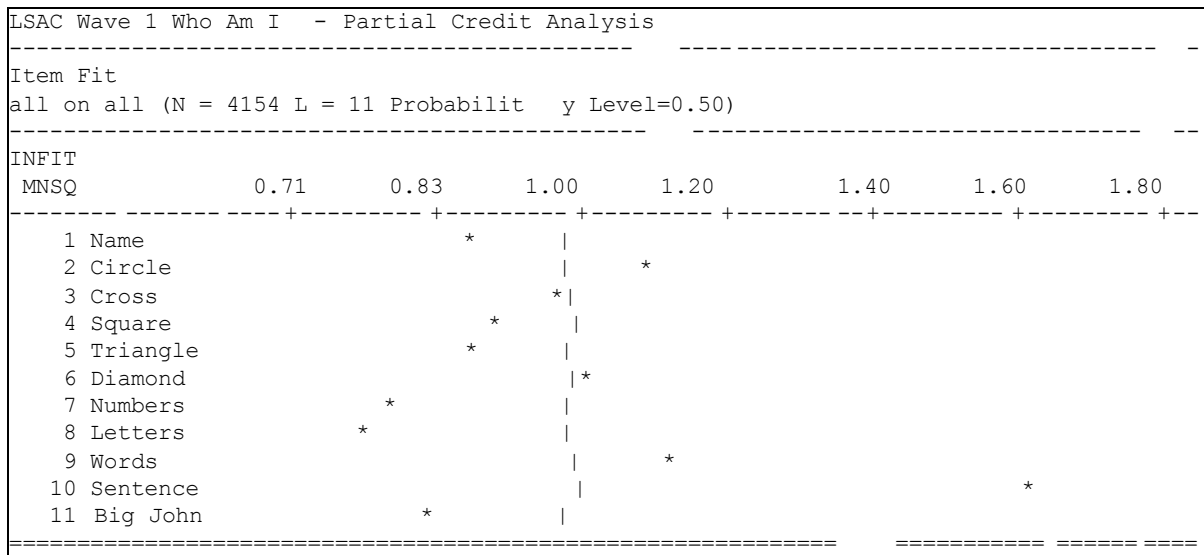
***Who Am I?***

*Who Am I?* consisted of 11 pages on which children were to write their names, copy shapes and write words and numbers. Each response was assessed on a four-point scale relating to the skill required for the task. A score of 0 was assigned if no attempt was made on the item. The data were fit using a partial credit item response model. The final case estimates were transformed to a scale with a mean of 64 and standard deviation of 8. Summary statistics are shown in Table 4.

In Wave 1, 4827 children were administered *Who Am I?* The test had a person separation reliability of 0.89.

*Quality of the test*

The statistics indicate that the *Who Am I?* data fit the rating scale model well, with most items falling within acceptable ranges, as shown in Figure 8. The most difficult item on the test was item 10, in which children were asked to write a sentence. Only 9 children received 4 points for their response; more than one-half of children made no attempt on this item. This is also shown in the item fit map (Figure 8): item 10 (Sentence) has an infit mean square of 1.67, while all other items ranged from 0.77 to 1.14.



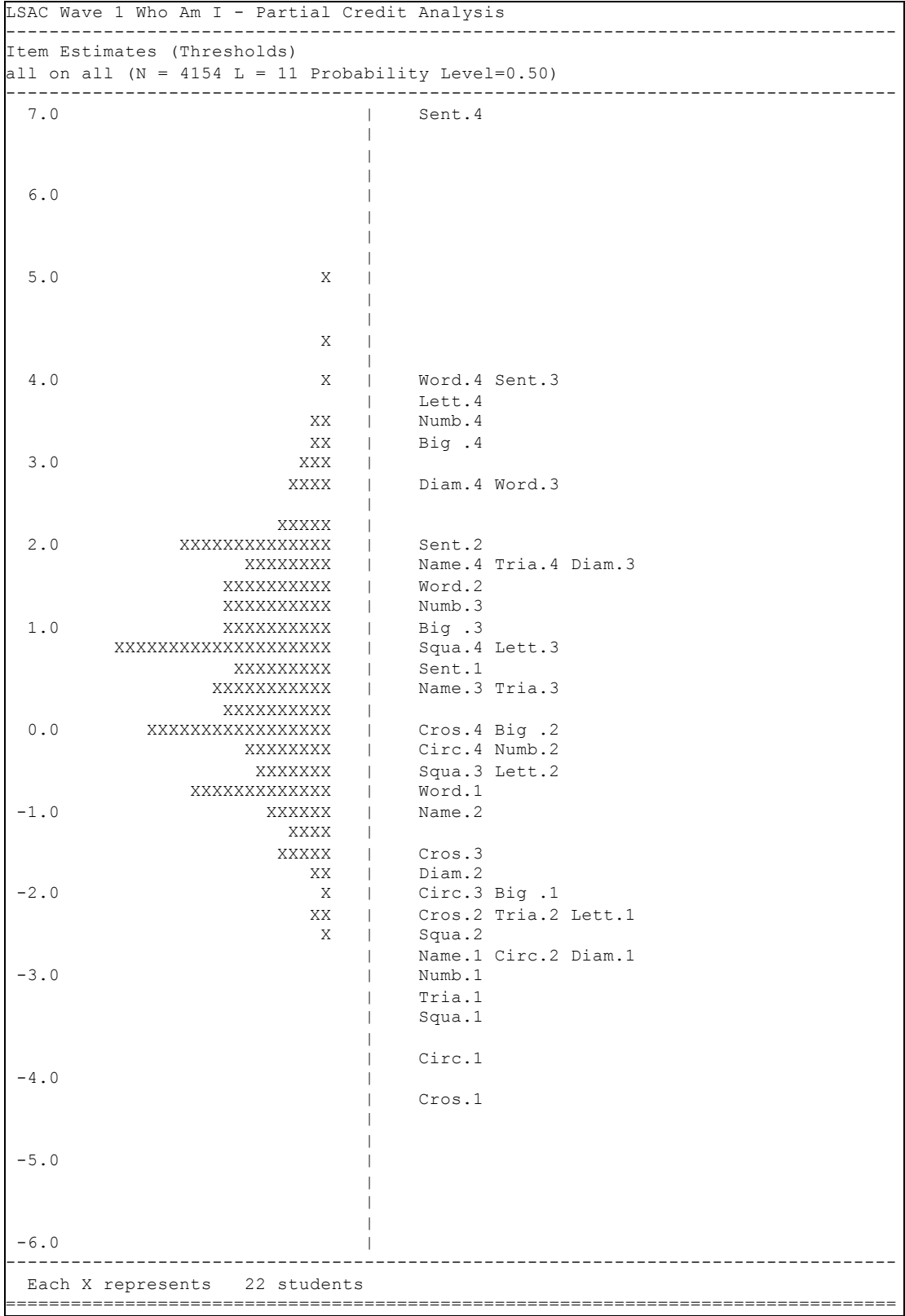
**Figure 2. Item fit map for all items on the *Who Am I?* test.**

Analysis of a partial credit model provides information on the thresholds required to move from any score to a higher score on each item. This information is provided in Figure 6, the item map, and is plotted against the distribution of case estimates. For all items, higher scores had higher logit values, indicating that higher scores were achieved by children who had higher ability estimates. The item map also indicates that the distribution of children (on the left) was only slightly above the distribution of the items (on the right), indicating that there was a good match between the difficulty of the items and the children’s ability levels.

**Correlation between the Adapted PPVT-III and *Who Am I?***

The two tests had a Pearson product-moment correlation of 0.309, based on 4386 children who received scores on both tests. While significant at the .01 level, this is considered a low-to-moderate correlation.

Sheldon Rothman  
ACER, 16 March 2005



Figure

re 3. Item map for all cases on the *Who Am I?* test.

## ***Part B: Wave 2 PPVT Development***

This paper describes the procedures used to develop a shortened version of the Peabody Picture Vocabulary Test (PPVT) for use in the Longitudinal Study of Australian Children (LSAC), *Growing Up in Australia*. This version of the test is to be used with six year-olds as part of the second wave of LSAC and is linked to the shortened version developed in 2003 for use with four year-olds as part of the first wave of LSAC.<sup>2</sup> The concept of a shortened version of the PPVT is from work done in the United States for the Head Start Impact Study. The original work was described in a draft paper by Philip Fletcher of Westat.

### ***Procedures***

All procedures described below are based on procedures used for the version used with four year-olds. For that version, four alternative sets of items were tested; one set was developed for use in LSAC. For the six year-old version, no alternative sets were used as the scope of the project was to develop a test that could be linked to the four year-old version.

As done for the test administered to four year-olds, the purpose was to develop a test that would consist of 40 items divided into a core set of 20 items, a basal set of 10 items for children who miss a minimum number of items on the core set, and a ceiling set of 10 items for children who correctly answer a minimum number of items on the core set. No child would take more than 30 items. It was also decided that at least 50 per cent of children should be required to take the core set only.

### **Testing**

A sample comprising 421 children was drawn from schools in New South Wales, Victoria and Queensland. During July and August 2005 the children were administered the full version of the PPVT-III, Form A, using the standard procedures for administering the test to six year-olds. These children ranged in age from 5 years 7 months to 7 years 11 months. Seventy-eight per cent of the children were six years old, and eighteen per cent were 7 years old. All children were in the same classes at the schools involved in the data collection. Subsequent examination of the data showed that the children from out-of-range ages did not appear as significantly different cases.

### **Analysis**

Test items were examined using a one-parameter logistic IRT model with the software Quest. For items below the PPVT basal set that were not administered, all were marked as correct. Items with a low number of responses were eliminated from the IRT analysis.

---

<sup>2</sup> Sheldon Rothman, 'An Australian Version of the Adaptive PPVT-III for Use in Research', unpublished paper (Melbourne: Australian Council for Educational Research, June 2003).

Overall, 132 items were used for analysis, as they covered a range that would allow 40 items to be selected and included the items administered in the four year-old test.

### **Selection of items**

The properties of the items were then determined, based on the data available from the Quest output. The first stage was to identify link items from the four year-old test that could be used with the six year-olds. For the 20 items of the core set, 8 items that had appeared in the four year-old test were selected. These items were selected on the basis of infit mean square and outfit mean square close to 1.00 in both administrations, the degree of difficulty on the items among both groups, the consistency of change between the administration to the groups, and the ability to provide a reasonable spread across the core set. Two items from the four year-old test were selected for the basal set, and one item from the four year-old test was selected for the ceiling set.

After the link items were selected, the remaining items were selected using those with infit mean square and outfit mean square close to 1.00, good discrimination and an ability to provide a reasonable range of item difficulties (-2.50 to +2.50). Items were also selected according to their position in the original PPVT sets and their parts of speech: nouns, verbs and adjectives. The final 20 core items were then positioned into two sets of 10 items, with the first 10 items generally easier than the second 10 items, but with an overlap of item estimates. Similar analyses were done to select the 10 basal and 10 ceiling items.

**Table 2. Items selected for Adaptive PPVT-III for use with 6 year-olds in LSAC.**

<b>Set</b>	<b>PPVT-III Form A item number</b>	<b>Item</b>	<b>Item threshold</b>	<b>Infit mean square</b>
<b>Core 1</b>	42	harp*	-2.55	1.01
	74	nostril*	-2.29	0.96
	56	furry*	-2.08	0.96
	52	diving*	-1.99	1.02
	78	horrified*	-1.44	0.99
	67	calculator	-0.38	1.10
	77	towing	-0.12	1.02
	91	clarinet	-0.02	1.07
	107	fern	0.53	1.03
	118	archery	0.88	0.98
<b>Core 2</b>	66	swamp*	-0.47	1.13
	90	interviewing*	-0.20	1.00
	96	vine*	0.10	0.97
	88	surprised	0.61	1.02
	68	signal	0.91	1.03
	114	injecting	0.97	0.99
	128	wailing	1.29	0.94
	131	foundation	1.85	0.98
	140	pastry	2.33	0.99
	125	valve	2.74	0.98
<b>Basal</b>	45	juggling	-4.98	0.74
	32	fountain*	-3.85	0.97
	40	farm*	-3.26	0.99
	47	tearing	-2.98	0.77
	49	parachute	-2.19	0.93
	71	vegetable	-1.70	1.04
	57	drilling	-1.62	0.92
	61	vehicle	-1.30	0.99
	75	vase	-1.21	0.94
	85	flamingo	-0.52	0.97
<b>Ceiling</b>	122	dilapidated*	1.11	0.98
	97	pedal	1.85	1.03
	149	abrasive	1.97	1.09
	143	pedestrian	2.07	0.97
	117	microscope	2.15	1.07
	153	detonation	2.69	0.94
	151	cascade	2.96	0.91
	139	consuming	3.57	1.04
148	replenishing	4.58	1.14	
	167	talon	--	--

\*Link item included in test for 4 year-olds

Note: Item threshold and infit mean square statistics are from the simulated test.

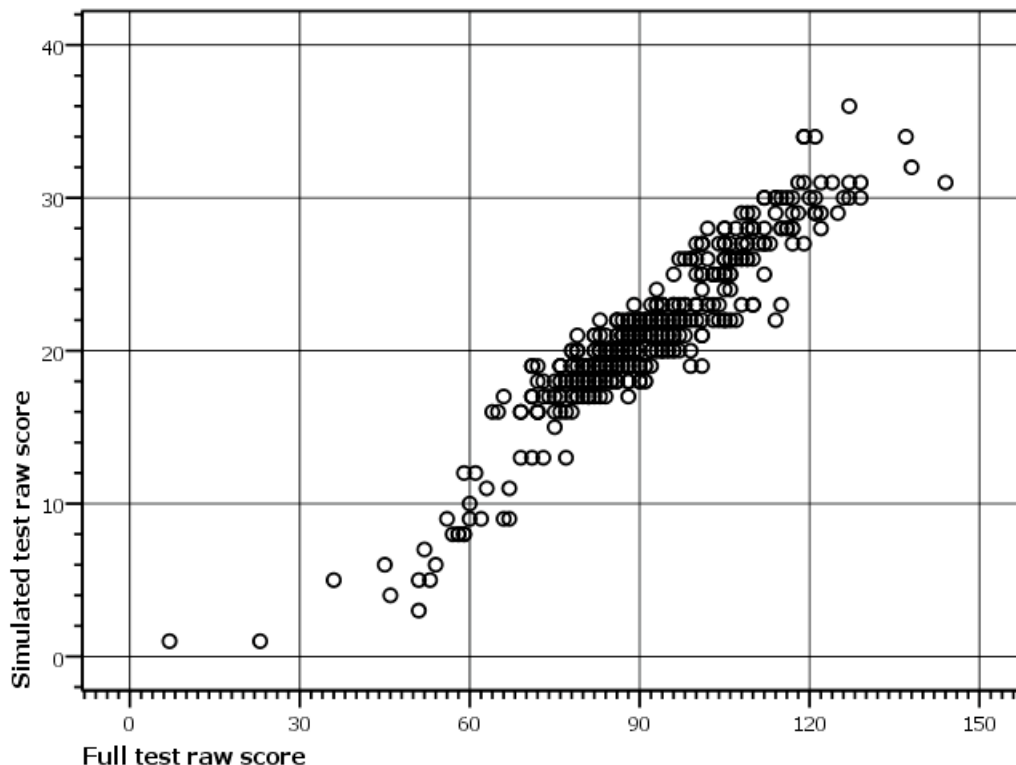


## Simulations

Simulation results were then calculated in SPSS. With the objective of having approximately 50 per cent of children requiring only the core set of 20 items, the items were checked to determine percentages of children requiring the basal or ceiling sets. Rules governing the administration of the test, particularly those regarding the number of correct items required for administration of the basal or ceiling sets, also guided the selection of items. The simulation suggests that approximately 25 per cent of children will require the ceiling set, 10 per cent the basal set and 65 per cent the core only. It should be noted that similar targets for the four year-olds were not achieved in the first wave of LSAC, with more than 65 per cent of children requiring the ceiling set.

Once the 40 items were selected, a new IRT analysis was conducted using only those items. Case and item estimates showed that the model fit the data well.

Results for the two versions—the full PPVT and the adapted version—were then compared: the full PPVT raw score with the case estimates from both the full test and the adapted test. The lowest correlation was 0.887; most correlations were in the 0.93 to 0.97 range, suggesting that the adapted version of the PPVT provides similar results to the full PPVT.



**Figure 4. Scatterplot showing joint distribution of scores on simulated adaptive PPVT-III and scores on full PPVT-III for 6 year-olds.**

## **Comment**

The six year-old version of the adaptive PPVT-III for use in LSAY was developed so that it can be linked with the four year-old version of the test. This allows the measurement of growth between administrations of the test. Forty items were selected for the test, with 20 items administered to all children as the core set (core sets 1 and 2). For children who make 15 to 20 errors an additional basal set of 10 items is administered, and for children who make 0 to 6 errors an additional ceiling set of 10 items is administered. Simulations showed that approximately two-thirds (67%) of children would require only the core sets, 7 per cent would require the core and basal sets, and 26 per cent would require the core and ceiling sets; this distribution was considered in the selection of items.

*Sheldon Rothman  
Principal Research Fellow  
9 November 2005*