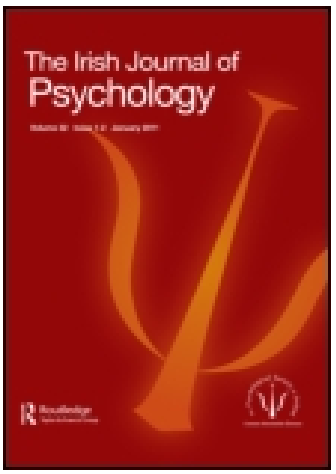

The Irish Journal of Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/riri20

Massive IQ Gains on the Scottish WISC: Evidence Against Brand et al.'s Hypothesis

James R. Flynn [a]

[a] University of Otago , New Zealand

Published online: 13 Nov 2012.

To cite this article: James R. Flynn (1990) Massive IQ Gains on the Scottish WISC: Evidence Against Brand et al.'s Hypothesis, The Irish Journal of Psychology, 11:1, 41-51, DOI: 10.1080/03033910.1990.10557787

To link to this article: http://dx.doi.org/10.1080/03033910.1990.10557787

The Irish Journal of Psychology, 1990, 11,1, 41—51

Massive IQ Gains on the Scottish WISC: Evidence Against Brand et al.'s Hypothesis

James R. Flynn

University of Otago, New Zealand

Fourteen nations have enjoyed massive IQ gains over the last generation, between 1950 and 1980, averaging 18 points on culturally reduced tests, 15 points on Wechsler-Binet tests and 11 points on purely verbal tests (Flynn, 1987b). Brand et al. (1989) hypothesised that these gains might be caused by altered test-taking techniques over time, techniques which confer an advantage on time-limited, multiple choice tests. As a test of their hypothesis, they predicted that IQ gains would be minimal on Wechsler verbal subtests because these eschew both multiple-choice and time pressure. They argued that Scottish WISC data confirm this prediction. However, it is argued here that those data show massive gains, count against their hypothesis, and support the conclusion that IQ tests cannot measure intelligence trends over time.

Scottish data from the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Intelligence Scale for Children revised (WISC-R) suggest massive IQ gains, rather than minimal gains as Brand, Freshwater and Dockrell (1989) have claimed. Their estimate rests on a methodological mistake. Here I will attempt to do four things: describe their mistake; offer indirect evidence of massive gains; provide an estimate of Scottish WISC gains; and draw implications for IQ tests as measures of intelligence trends over time.

The methodological mistake

The Scottish WISC-R standardisation sample of 1983/4 outperformed the Scottish WISC sample of 1961/2 on 68 items that were left unaltered between the two tests. Data exist only for 10 and 13-year-olds. Brand et al. estimated Scottish IQ gains over that period by calculating the typical rise in the percentage passing (giving correct or semi-correct answers) each of those 68

Requests for reprints should be addressed to Professor James R. Flynn, Department of Political Studies, University of Otago, P.O. Box 56, Dunedin, New Zealand.

items. Without explanation, they did this by using a series of medians rather than simply calculating the average increase in the percentage passing. However, my main point is this: they took the typical rise, calculated item by item, as a measure of overall improved performance on the test as a whole. This method would work only if the correlation between each item of a mental test and the whole test were perfect, which we all know to be false. If the correlation is, for example, .154, then simply getting the improvement item by item ignores a huge regression effect, and the estimate of the overall improvement would only be .154 of the true value!

Let me illustrate: assume true enhanced performance for Wechsler verbal IQ was 10 points. Each of the five verbal subtests correlate with verbal IQ at only .70, so if you averaged the subtest scores you would get an estimate of only 7 points (10 x .70 = 7). Taking the subtests one by one, I have calculated a correlation of .22 between a typical item and a typical subtest. So if you calculated the improvement on a given subtest by averaging item by item scores, you would reduce your estimate to 1.54 points (7 x .22 = 1.54). Moreover, it is a pecularity of these data that medians give a lower value than averages — about .55 as much. So using medians would reduce the estimate to .847 points (1.54 x .55 = .847). It is for reasons like this that Brand et al. could estimate gains at "less than 1 IQ point" when the true gains are about 10 points. By selecting items with maximum discrimination, they boosted their estimate to almost 2 points, or 2.5 points when projected over thirty years. However, that too is a gross underestimate.

The validity of the above can be verified simply by reducing all gains to standard deviation units and using the WISC scoring tables. For example, assume for each verbal subtest a scaled score gain of 1.4 points. As the *SD* for each subtest is set at 3 points, this amounts to .47*SD* (1.4/3 =.47). Over five verbal subtests, the total gain would be 7 scaled score points (1.4 x 5 = 7). Using the WISC conversion table for the verbal scale, a scaled score of 57 (7 points above average) translates into an IQ of 110 (10 points above average) and 10 IQ points amounts to .67 *SD*. The gain has grown from .47 *SD*, subtest by subtest, to .67 *SD* on the verbal scale as a whole, because the correlation between a typical subtest and the whole is only .70 (.47/.70 = .67).

Indirect evidence of massive gains

Two kinds of indirect evidence support regression analysis in evidencing that Scottish IQ gains have been massive. When the 1983/4 sample is scored against the 1961/2 sample, it has a mental age far above its chronological age. When Scottish standardisation samples are scored against white American samples, their gains are seen to exceed massive American gains.

The mental age comparison involves four steps:

1. From the Scottish WISC manual, calculating the mean raw scores for all ages from 10 years upward on the total collection of 116 verbal items (Wechsler, 1965, pp. 32, 35, 38, 41, 44 and 47).
2. Calculating the mean raw scores for the 10-year-olds and 13-year-olds from the 1961/2 WISC sample on the 56 verbal items that were left unaltered; the 12 performance items will be set aside for the moment. The number of correct answers for each verbal subtest can be calculated simply by adding together the percentages passing each item and dividing the total by 100. For the Comprehension, Similarities and Vocabulary subtests, you must then multiply by 2 because two raw score points are given for each fully correct answer.
3. Calculating the mean raw scores for other ages from the 1961/2 WISC sample on the 56 unaltered items. This is done on the assumption that the missing scores would mimic the pattern of increase by age set by raw scores on the total collection of 116 items.
4. Calculating the mean raw scores of the 10-year-olds and 13-year-olds from the 1983/4 sample on the 56 unaltered items. The method is identical with the same calculation for the earlier sample.

Table 1. Verbal IQs (based on mental age) of the 1983/4 sample when scored against 1961/2 norms.

Sample	Number of verbal items	Perfect score[a]	Actual score by age[b]					
			10	11	12	13	14	15
1961/2	116	182	79.0	85.5	92.0	100.0	104.5	112.0
1961/2	56	93	42.7	45.7	48.6	52.3	54.3	57.7
1983/4	56	93	47.6	--	--	57.7	--	--

Sample	Age		Verbal IQ	Calculation
	Mental[c]	Chronological		
1983/4	11.66	10.00	117	(11.66/10.00)100=117
1983/4	15.00	13.00	115	(15.00/13.00)100=115

Note. [a] The perfect score is the raw score obtained when every item on the WISC verbal scale is scored correct; some items give two points for a fully correct answer. [b] The actual score is the mean raw score for each age. [c] The mental age of the 1983/4 sample shows what chronological age from the 1961/2 sample had the same mean raw score (e.g., 10-year-olds from 1983/4 matched the raw score of 11.66-year-olds from 1961/2).

These calculations are set out in Table 1, which compares the two samples on the 56 unaltered verbal items, without speculating about how they would have done on the items that were replaced. The result: the 10-year-olds of 1983/4 matched the raw scores of the 11.66-year-olds of 1961/2, and the 13-year-olds of 1983/4 matched the raw scores of the 15-year-olds of 1961/2. The usual method of calculating mental age IQs gives the later 10-year-olds a mean of 117 when scored against the earlier norms and the later 13-year-olds a mean of 115, signalling massive IQ gains over a period of 22 years. As we will see, the later sample had a slight age advantage over the earlier one, which would lower these values by about two points. Adding in the data from Picture Completion, the sole performance subtest with unaltered items, makes a negligible difference: none to the 10-year-olds, but it cuts an extra point off the 13-year-olds. I should point out that even adjusted values — means of 115 and 112 — would still be too high, because mental age IQs tend to be inflated compared with the deviation IQs the Scottish WISC uses, and later we will get a proper estimate using the WISC scoring tables. Nonetheless, the mental age data are highly significant for the argument: when subjects have a mental age one or two years beyond their chronological age, there can be no plausible case for claiming that their IQ advantage is negligible.

Another piece of evidence: the Scottish Council for Research in Education (SCRE) has compared its own standardisation samples with the relevent American samples. It selected 93 subjects aged 6 to 15 years from the Scottish WISC sample and used their raw scores to score them against both the Scottish and the American norms. The results showed a Scottish IQ deficit, that is, the Scottish WISC sample had performed below the American WISC sample and established weaker norms (SCRE, 1967, p. 26)[1]. The results must be treated with caution: the number of subjects at each age was too few to give age-specific data and the American test was adapted in minor ways to suit Scottish conditions. Some 20 years later the SCRE scored its entire WISC-R sample against American norms, and this time the results showed a Scottish IQ credit, that is, their scores were consistently above 100 (SCRE, 1987, p. 23). Finally, there are the results of comparative studies within America: 33 samples totalling 1607 subjects took both the WISC and the WISC-R and they show that white Americans from the WISC-R sample set norms that were 8.54 points above those set by the all-white WISC sample. This represents the difference for fullscale IQ, the verbal difference was about a point less, that for performance a point more (Flynn, 1984, p. 283; 1985, Table 1).

The pattern is clear: Scottish children went from below the American WISC

[1] Note the typographical error which makes the reverse seem true for the verbal results.

sample to above the American WISC-R sample, despite the fact that the latter outperformed the former by a huge margin. To estimate Scottish IQ gains between 1961/2 and 1983/4, all we need do is take the Scottish WISC deficit, add on the American WISC to WISC-R gain, and add on the Scottish WISC-R credit. For obvious reasons, I have calculated the Scottish WISC-R credit using the 10 and 13-year-olds, although their scores, when averaged, are almost identical with all ages; and I have adjusted their scores to make them comparable to white Americans only, which essentially meant deducting about 2 points from each score.

The necessary arithmetic is shown in Table 2; it shows Scottish gains of approximately 9 points for verbal IQ, 15 points for performance, and 12 points for fullscale. However, let us leave indirect and rough estimates of Scottish gains for a more precise estimate.

Table 2. Scottish IQ gains from 1961/2 to 1983/4 (based on comparisons between Scottish and white American standardisation samples).

	Samples compared			
IQ	Scottish WISC scored against American WISC	American WISC-R scored against American WISC[a]	Scottish WISC-R scored against American WISC-R	Scottish IQ gain
Verbal	0.70 (deficit)	7.54 (credit)	0.45 (credit)	8.69
Performance	3.36 (deficit)	9.54 (credit)	2.58 (credit)	15.48
Fullscale	1.45 (deficit)	8.54 (credit)	1.65 (credit)	11.64

Notes. The Scottish WISC sample consisted of 93 subjects selected from those aged 6 - 15 years. The Scottish WISC-R sample consisted of 296 subjects aged 10 and 13 years. [a] Based on 1607 subjects who took both tests.

Scottish IQ gains from 1961/2 to 1983/4

The best method of calculating these gains is to score the 10 and 13-year-olds from the Scottish WISC-R sample directly against the WISC manual, using the 56 verbal items that were unaltered from the WISC to WISC-R as our data base. This involves six steps:

1. For each verbal subtest, calculate the number of unaltered items the WISC and WISC-R samples each got correct. Using Comprehension as an example, adding together the percentages who passed each item and dividing by 100, the WISC-R 10-year-olds got 4.57 of the nine items correct, the WISC subjects 3.78. Fully correct answers get 2 points, so the raw scores are 9.14 and 7.56, respectively.

2. Get the ratio between the raw scores of the two samples, so 9.14 divided by 7.56 gives a ratio of 1.21:1 in favour of the WISC-R subjects.
3. Adjust this ratio so that it is based on results from items that simulate sampling the full range of difficulty of WISC items. The nine unaltered items from Comprehension are close to that, but still it is one easy item short, so a dummy item has to be created that matches the averaged results from the four existing, easy items. Over these 10 items, the raw scores were 10.79 and 9.04 respectively, yielding a ratio of 1.19:1. I will give both adjusted and unadjusted ratios for all subtests, so the reader can see that no skullduggery is being perpetrated. In fact, adjusted ratios give slightly lower verbal IQ gains: their main value is more accurate results subtest by subtest.
4. With a ratio based on items that simulate a random selection from all 28 Comprehension items, we simply multiply it by the WISC sample raw score for Comprehension given in the manual, and we get what the WISC-R sample would have got as a raw score. Here, there is a complication: the WISC 10-year-olds were only 10 years 3.9 months rather than the 10 years and 6 months they should have been (SCRE, 1967, p.19). The WISC-R subjects were the correct age (SCRE, personal communication, 3 August, 1988) and being a bit older, had an unfair advantage. To compensate, the ratio in their favour must be multiplied by the raw score appropiate for the actual age of the younger WISC subjects: which means pro-rating the values from the two relevant WISC tables (Wechsler, 1965, pp. 31 and 32). For Comprehension, both tables give 12.50 as the raw score, but for other subtests the tables give a lower and a higher value, and pro-rating provides something in between. We now have our key result: 1.19 (ratio) times 12.50 (WISC raw score) = 14.88 as the WISC-R 10-year-old raw score on Comprehension. As for the WISC 13-year-olds, they were really only 13 years 2.7 months, so even more radical pro-rating was done to calculate appropriate raw scores for them (Wechsler, 1965, pp. 40 and 41).
5. The WISC-R 10-year-old raw score can now be used to get their scaled score from the WISC table that matches their age (Wechsler, 1965, p. 32). A raw score of 14.88 on Comprehension gives a scaled score of 12.88. Using the subtest mean of 10 and *SD* of 3, we can calculate their IQ on Comprehenison scored against WISC norms: clearly they are almost a full standard deviation above the mean and earn an IQ of 114.40.
6. To get their total verbal IQ, we merely sum the scaled scores from the five verbal subtests — this total is 56.88. Using the manual's conversion table (Wechsler, 1965, p. 14), that gives a verbal IQ of 109.88. Therefore, the true verbal IQ gain of Scottish 10-year-olds from 1961 (the actual year they were tested, see SCRE, 1967, p. 19) to 1983/4 was 9.88 points over 22.5

years. This gives a rate of gain of .439 points per year and projected over 30 years, a generational gain of 13.17 points. The 13-year-olds were tested in 1962 (SCRE, 1967, p. 19) and gained 9.55 points over 21.5 years, which gives a rate of .444 points per year and a 30-year generational gain of 13.33 points.

All the relevant values for all five WISC verbal subtests are given in Table 3. As promised, it also gives ratios derived from the 56 unaltered items alone; those who prefer to do so can use them and ignore the dummy items created. Comparing Table 3 with Table 2, verbal gains of 9 to 10 points from precise estimates are remarkably close to the rough 9-point estimate based on American comparisons. This has emboldened me to carry over the performance

Table 3. Scottish IQ gains on Wechsler subtests and the total Verbal scale (based on scoring the 1983/4 WISC-R sample against 1961/2 WISC norms).

Sub-test	Number of items[a]	Raw scores 1983/4	1961/2	Ratio	1961/2 mean	1983/4 mean	Scaled score	IQ
				10-year-olds				
I	21 (14)	9.74	8.77	1.11 (1.02)	12.48	13.85	10.85	104.25
C	10 (9)	10.79	9.04	1.19 (1.21)	12.50	14.88	12.88	114.40
A	9 (5)	5.32	5.51	0.97 (1.00)	9.74	9.45	9.45	97.25
V	26 (21)	22.33	20.92	1.07 (1.06)	31.98	34.21	10.00	100.00
S	10 (7)	6.92	3.86	1.77 (1.88)	5.48	13.70[b]	13.70	118.50
Total	76 (56)						56.88	109.88
				13-year-olds				
I	23 (14)	14.19	13.14	1.08 (1.04)	17.18	18.55	10.55	102.75
C	10 (9)	13.79	11.33	1.22 (1.24)	16.00	19.52	13.52	117.60
A	15 (5)	10.13	11.22	0.90 (0.98)	12.00	10.80	8.80	94.00
V	26 (21)	26.64	25.91	1.03 (1.03)	40.18	41.38	10.00	100.00
S	11 (7)	12.60	8.30	1.52 (1.64)	9.00	17.68[b]	13.68	118.40
Total	85 (56)						56.55	109.55

Note. I—Information; C—Comprehension; A—Arithmetic; V—Vocabulary; S—Similarities. [a] The numbers in brackets refer to the actual items unaltered from the WISC to the WISC-R; the additional items were created as described in the text. See text for explanations of other terms .[b] For Similarities, 4 points are credited for 'too easy' items and these must be added to get the 1983/4 mean.

Table 4. Summary of Scottish IQ gains on Wechsler tests.

Test	Actual Gain (IQ points)	Actual Period (years)	Rate of gain (points/year)	Projected 30-year gain (IQ points)
		10-year-olds		
Verbal	9.88	22.5	.439	13.17
Performance	15.83	22.5	.704	21.11
Fullscale	12.86	22.5	.571	17.14
		13-year-olds		
Verbal	9.55	21.5	.444	13.33
Performance	15.13	21.5	.704	21.11
Fullscale	12.34	21.5	.574	17.22

gain based on those comparisons so Table 4 offers a full range of estimates for Scottish IQ gains on Wechsler tests. Projected over thirty years, they are 13 points for verbal IQ, 21 points for performance IQ and 17 points for fullscale IQ (here simply the average of the other two). There is no appreciable difference between 10 and 13-year-olds, but note that performance gains have been defined as identical because the relevant data were not age specific.

Implications

Brand et al. (1989) developed a complex causal hypothesis along the following lines: the advent of the permissive society over the last fifty years may have caused a characterological change towards personal liberalism. The growth of personal liberalism may have caused a change in test-taking technique by promoting a tendency towards quick, intuitive responses or intelligent guessing. This technique advantages current youth on time-limited multiple-choice tests, particularly culturally reduced tests, while the scrupulous and painstaking child of the past was disadvantaged by exhausting his or her time and energy trying to get every item correct. This hypothesis is purported to explain massive generational IQ gains, for example, an average gain of 18 points on Raven's Progressive Matrices over fourteen nations (Flynn, 1987b). Brand (1987b) suggested that such IQ gains are correlated with a variety of social trends, increasing rates of sexual promiscuity, illegitimacy, divorce, irreligiosity, cigarette consumption, accidents, crime and

Britain's world leadership in the field of popular music. This hypothesis is also held to justify a choice of less defective tests as measures of intelligence. Wechsler IQ tests are recommended, particularly the verbal scales, as largely eschewing multiple-choice and exerting little time pressure because of their mode of individual administration.

Therefore, Brand et al. (1989) selected Scottish trends from the WISC to the WISC-R for verbal IQ as a test of their hypothesis, with emphasis not only on the character of the test but on the excellence of the samples. If I am correct, the Scottish WISC to WISC-R data show a massive 13-point generational gain for verbal IQ and, therefore, these data count against Brand et al.'s hypothesis. It may be said that data from West Germany, Austria and Japan had already cast doubt on their hypothesis: these nations have registered verbal gains ranging from 12 to 20 points per generation (Flynn, 1987b, p.186). But perhaps Brand et al. considered the Scottish samples to be the best existent. Different scholars have every right to invest their time in whatever hypothesis they prefer, but in my opinion, the drift of evidence is such that other causal explanations of this baffling phenomenon should take precedence (Brand, 1987a; Flynn, 1987a; Flynn, in press).

The Scottish WISC to WISC-R data also add to that body of data which casts doubt on another popular hypothesis. I refer to Lynn's speculation that massive IQ gains occur when earlier samples were born during either the Great Depression or World War 2, thereby suffering hardships, and later samples were born post-war, thereby enjoying more favourable circumstances (Lynn, 1987: Flynn, 1988, p.349). The Scottish WISC 10-year-olds were of course born in 1951, the 13-year-olds in 1949.

The Scottish data show different gains for different verbal subtests, enormous gains on Comprehension and Similarities, significant gains on Information, no gain on Vocabulary. The loss on Arithmetic should not be taken too seriously: as the large number of dummy items created shows, the five unaltered items were too few and far too easy to measure trends over time. Upon my request, Jensen (personal communication, 20 June, 1988), ranked the WISC verbal subtests in terms of school influence and while emphasizing the subjectivity of his assessment, classed Comprehension and Similarities as least influenced, Information, Arithmetic and Vocabulary as most so. Dockrell (personal communication, 20 July, 1988) gave a very similar ranking, and as a consequence was surprised by the gains on Comprehension and Similarities because he assumed that whatever gains occurred would be caused primarily by enhanced academic achievement over time. In other words, the Scottish data show exactly what the data from fourteen nations showed: IQ gains increase the further we go from school taught skills and content towards whatever IQ tests measure beyond academic achievement (Flynn, 1987b, pp. 184-188). In addition, they suggest that altered test-taking technique has not been a factor.

The fourteen nation's data posed an important problem for IQ tests: culturally reduced tests, Wechsler performance IQ and Wechsler verbal IQ all gave generational gains far too large to be identified with intelligence gains, which meant they could not plausibly be used to measure intelligence trends over time. Now we are reduced to trying to salvage particular Wechsler subtests. The Scottish data show Arithmetic, Vocabulary and perhaps Information as still viable, but West German and Austrian data show large gains on Information, enormous gains on Vocabulary and, incidentally, enormous gains on all Wechsler performance subtests (Schallberger, 1985, Table 4; Schubert & Berlach, 1982, Table 2). And no one would really use Arithmetic in isolation as an intelligence test. To measure trends over time, we are left with nothing better than our subjective assessment that human beings, living their every-day lives, have shown pretty much the same mixture of intelligence and stupidity throughout recorded history.

REFERENCES

Brand, C. R. (1987a). Bryter still and bryter? *Nature*, **328**, 110.

Brand, C. R. (1987b). Keeping up with the times. *Nature*, **328**,761.

Brand, C. R., Freshwater, S. & Dockrell, W. B. (1989). Has there been a 'massive' rise in IQ levels in the West? Evidence from Scottish children. *The Irish Journal of Psychology*, **10**, 388-394.

Flynn, J. R. (1984). IQ gains and the Binet decrements. *Journal of Educational Measurement*, **21**, 283-290.

Flynn, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency*, **90**, 236-244.

Flynn, J. R. (1987a). Causal factors in generational IQ gains. *Nature*, **328**, 765.

Flynn, J. R. (1987b). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, **101**, 171-191.

Flynn, J. R. (1988). Japanese intelligence simply fades away: A rejoinder to Lynn. *The Psychologist*, **1**, 348-350.

Flynn, J. R. (in press). Cultural distance and the limitations of IQ. In J. Lynch, C. Modgil & S. Modgil (Eds), *Education for Cultural Diversity: Convergence and Divergence*. Lewes, Sussex: Falmer Press.

Lynn, R. (1987). Japan: Land of the rising IQ. A reply to Flynn. *Bulletin of the British Psychological Society*, **40**, 464-468.

Schallberger, U. (1985). *HAWIK und HAWIK-R: Ein empirischer Vergleich* [HAWIK and HAWIK-R: an empirical comparison] (Technical Report). Zürich: Psychologisches Institut der Universität.

Schubert, M. T. & Berlach, G. (1982). Neue Richtlinien zur Interpretation des Hamburg Wechsler-Intelligenztests für Kinder (HAWIK). [New guidelines for the interpretation of the Hamburg Wechsler Intelligence Tests for Children (HAWIK)]. *Zeitschrift für Klinische Psychologie*, **11**, 253-279.

Scottish Council for Research in Education (1967). *The Scottish Standardisation of the Wechsler Intelligence Scale For Children*. London: University Press.

Scottish Council for Research in Education, (1987). *The Wechsler Intelligence Scale For Children — Revised: Scottish Standardisation*. London: The Psychological Corporation.

Wechsler, D. (1965). *The Wechsler Intelligence Scale For Children: Manual For Scottish Standardisation*. Edinburgh: Scottish Council for Research in Education.

(Accepted for publication April, 1990)