Rosenthal, Robert, and Jacobson, Lenore. *Pygmalion in the Class-
   room.* New York: Holt, Rinehart and Winston, 1968.
   240 + xi pp. $3.95.

The enterprise which represents the core of this document, and
presumably the excuse for its publication, has received wide-
spread advance publicity. In spite of anything I can say, I am
sure it will become a classic—widely referred to and rarely ex-
amined critically. Alas, it is so defective technically that one can
only regret that it ever got beyond the eyes of the original in-
vestigators! Though the volume may be an effective addition to
educational propagandizing, it does nothing to raise the standards
of educational research.

   Though it may make for a dull review, I feel I must dissect the
study to point out some basic defects in its data that make its
conclusions (though they may possibly be true) in no sense ade-
quately supported by the data. The general reasonableness of the
"self-fulfilling prophecy effect" is not at issue, nor is the reported
background of previous anecdote, observation, and research. The
one point at which this review is directed is the adequacy of pro-
cedures (of data gathering and data analysis) and the appro-
priateness of the conclusions drawn from the study that consti-
tutes the middle third of the book.

   Before we can dig beneath the surface, we must outline briefly
on a surface level what was done and what was reportedly found.
In May 1964, the SRA-published Tests of General Ability
(TOGA) were administered by the classroom teachers to all pupils
in kindergarden and all six grades of a school. The test had been
presented to the teachers as a test that ". . . will allow us to predict
which youngsters are most likely to show an academic spurt."
The following September each teacher was given a list of names
of pupils (actually selected by a table of random numbers) who
were alleged to be the ones likely to show a spurt.

   The children were tested again in January 1965, May 1965,
and May 1966. The authors assert that the results support the
proposition that the teachers' expectancies influenced the mental
development of the children.

   The main results of testing in May 1965 (from the authors'
Table 7-1) are as follows:

	Control			Experimental		
Grade	*N*	*Gain*		*N*	*Gain*	*Difference*
1	48	+12.0		7	+27.4	15.4
2	47	+ 7.0		12	+16.5	9.5
3	40	+ 5.0		14	+ 5.0	0
4	49	+ 2.2		12	+ 5.6	3.4
5	26	+17.5		9	+17.4	-0.1
6	45	+10.7		11	+10.0	-0.7

Thus, to all intents and purposes, the alleged effect of the "proph-ecy" appears in 19 children in grades 1 and 2. If we are to trust the results, and the large edifice of further analysis and specula-tion built upon them, the findings for these two grades must be unimpeachable. Let us examine them.

TOGA has two subtests, one consisting of oral vocabulary and one of multi-mental ("which one doesn't belong") items. For the K-2 level of the test, the one used in the pretesting and posttest-ing of grades 1 and 2, the two parts of the test contain respective-ly 35 and 28 items. Let us look first at the pretest data for six classrooms, three tested in kindergarden and three in the first grade. The results, from Appendix Tables A-2 and A-3 were (ex-pressed in numbers that are always spoken of by the authors as "IQs"):

	Experimental				Control	
		Mean Verbal	*Mean Reasoning*		*Mean Verbal*	*Mean Reasoning*
Class	*N*	*"IQ"*	*"IQ"*	*N*	*"IQ"*	*"IQ"*
1A	3	102.00	84.67	19	119.47	91.32
1B	4	116.25	54.00	16	104.25	47.19
1C	2	67.50	53.50	19	95.68	30.79
2A	6	114.33	112.50	19	111.53	100.95
2B	3	103.67	102.33	16	96.50	80.56
2C	5	90.20	77.40	14	82.21	73.93

On the Reasoning Test, one class of 19 pupils is reported to have a mean "IQ" of 31! They just barely appear to make the grade as imbeciles! And yet these pretest data were used blithely by the authors without even a reference to these fantastic results!

If these pretest data show anything, they show that the testing was utterly worthless and meaningless. The means and standard deviations for the total first and second grade classes were (calculated by combining sub-groups):

	First Grade		Second Grade	
	Mean	*S. D.*	*Mean*	*S.D.*
Verbal	105.7	21.2	99.4	16.1
Reasoning	58.0	36.8	89.1	21.6

What kind of a test, or what kind of testing is it that gives a mean "IQ" of 58 for the total entering first grade of a rather run-of-the-mill school?

Unfortunately, nowhere in the whole volume do the authors give any data expressed in raw scores. Neither do they give the ages of their groups. So it takes a little impressionistic estimating to try to reconstruct the picture. However, it would not be far off to assume an average age of 6.0 for May of a kindergarden year. An "IQ" of 58 would then mean a "mental age" of 3.5. So we go to to the norms tables of TOGA to find the raw score that would correspond to a "mental age" of 3.5. Alas, the norms do not go down that far! It is not possible to tell what the authors did, but finding that a raw score of 8 corresponds to an "M.A." of 5.3, we can take a shot at extrapolating downward. We come out with a raw score of approximately 2! Random marking would give 5 or 6 right!

We can only conclude that the pretest on the so-called Reasoning Test at this age is worthless. And, in the words of a European colleague, "When the clock strikes thirteen, doubt is cast not only on the last stroke but also on all that have gone before."

Another look at one of the Appendix tables (A-6) shows that the 6 pupils in class 2A who had been picked to be "spurters" have a reported mean and standard deviation of posttest "IQ" of 150.17 and 40.17 respectively. This looks a little high! What does it mean in raw score terms? Again, we must turn detective with somewhat inadequate clues. Not knowing pupil ages, let us assume 7½ as probably on the low side for May in the second grade. An "IQ" of 150 implies, then, a mean "M.A." of 11¼. Back to our TOGA norms to find the corresponding raw score.

Alas, the highest entry is 10.0 for a raw score of 26! We must once more extrapolate, and the best we can do from the existing data is to get 28+. (Remember, there are only 28 items in this sub-test.) The mean of 6 represents a perfect score! But the standard deviation is 40 "IQ" points. What of those who fall above the mean?

When the clock strikes 14, we throw away the clock!

In conclusion, then, the indications are that the basic data upon which this structure has been raised are so untrustworthy that any conclusions based upon them must be suspect. The conclusions may be correct, but if so it must be considered a fortunate coincidence.

<div align="right">

Robert L. Thorndike
Teachers College
Columbia University

</div>