

The Lake Wobegon Effect Reversed: Commentary on “The Gesell Assessment: Psychometric Properties”

Samuel J. Meisels

The University of Michigan

By now the “Lake Wobegon Effect” has become famous in public education. Taken from a description of Garrison Keillor’s mythical community, “where the women are strong, the men are good-looking, and all of the children are above average,” it has become a watchword for how tests can be used to distort achievement. In 1988, John Cannell, a West Virginia pediatrician, became alarmed that many of the children in his practice who were not doing well in school were reported to be performing “at an average level” on state-administered standardized tests. Cannell undertook a national investigation of these tests. To his shock and dismay, he found that all states reporting statewide test scores ranked their children above the national average. Thus, the phrase, “Lake Wobegon Effect.”

The problems uncovered by Cannell, and later largely confirmed by more systematic research, are brought to mind in this context. However, in the present situation, Walker reports a reverse Lake Wobegon effect: rather than finding that all children are above average, this study finds that children’s average performance on the *Gesell Developmental Assessment (GDA)* fell below chronological age expectations. The discrepancy ranged from 2 months at age 4 to nearly 7 months by age 6. Similarly, the average Grade Recommendations fell below those expected for the children’s chronological ages.

Findings of such magnitude and consistency would normally suggest that the assessment—the *GDA* in this case—is in need of recalibration, since in a representative sample it is unlikely that so many of the children would be delayed. However, Walker concludes that the problem lies within the children rather than the test. Lichtenstein (1990), reported similar findings for the full *Gesell School Readiness Screening Test (GSRST)*, of which the

Correspondence and requests for reprints should be sent to Samuel J. Meisels, Center for Human Growth and Development, University of Michigan, 300 North Ingalls, Ann Arbor, MI 48109-0406.

GDA is a prominent element. But unlike Walker, he interpreted the greater than 50% discrepancy between “developmental” and chronological age as evidence of the *GSRST*’s miscalibration.

In the present study, it could be objected that the *GDA*’s unexpected relationship to chronological age is vindicated by its correlations with follow-up assessments. Such correlations, if high enough, would indeed place the onus on the children rather than on the assessment. But before this conclusion can be drawn, two others must first be established. First, it must be shown that the correlations that were obtained were not influenced by teachers’ prior knowledge or exposure. Lichtenstein (1990) demonstrated that teachers’ “tendency to perceive children as unready is directly proportional to the extent of Gesell Institute training received” (p. 371). In other words, low scores on the *GDA* may have set up an expectancy among the children’s teachers concerning the children’s performance. Only a completely “blind” trial, in which the Gesell findings were concealed from the preschool and follow-up teachers, and the teachers were uninformed about Gesell teachings and practice, could eliminate this powerful source of potential bias.

Second, given that this bias is controlled, it must be demonstrated that the preschool indicator, the *GDA*, is highly predictive of the classifications obtained on the 8-year-old measures. Table 11 (p. 35) attempts to show this relationship, but in the vast majority of cases, children changed classifications in the follow-up assessment. Indeed, on three of the four outcome measures, the preschool ratings were lower than the average performance on that outcome across all quartiles (in one area, Reading, the prediction was identical to the outcome). Thus, it appears that the *GDA*’s underestimation of children’s abilities—the reverse Lake Wobegon effect—is highlighted by these comparisons. These data do not support the use of the *GDA* to place children into extra year programs.

In short, this study demonstrates the reliability of the *GDA*, but it leaves unresolved several key questions about its validity. Furthermore, through its consistent finding of developmental ratings below chronological age expectancies it asks us to believe in the test rather than in the child—a peculiar position indeed for advocates of developmentally appropriate practice.

REFERENCE

- Lichtenstein, R. (1990). Psychometric characteristics and appropriate use of the Gesell School Readiness Screening Test. *Early Childhood Research Quarterly*, 5, 359-378.