

American Educational Research Journal

<http://aerj.aera.net>

Do Teacher Ratings and Standardized Test Results of Students Yield the Same Information?

Joseph J. Pedulla, Peter W. Airasian and George F. Madaus

Am Educ Res J 1980 17: 303

DOI: 10.3102/00028312017003303

The online version of this article can be found at:

<http://aer.sagepub.com/content/17/3/303>

Published on behalf of



American Educational
Research Association

American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *American Educational Research Journal* can be found at:

Email Alerts: <http://aerj.aera.net/alerts>

Subscriptions: <http://aerj.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

Citations: <http://aer.sagepub.com/content/17/3/303.refs.html>

Do Teacher Ratings and Standardized Test Results of Students Yield the Same Information?

JOSEPH J. PEDULLA, PETER W. AIRASIAN and
GEORGE F. MADAUS

Boston College

Teachers in Ireland (n = 170) rated students (n = 2,617) on IQ, mathematics, and English, as well as on 12 social and academic classroom behaviors. Factor analysis of IQ, mathematics, and English standardized test scores, together with the 15 teacher ratings, showed that there is overlap between ratings and test results but that the information obtained is not redundant. Three factors were identified: one was comprised primarily of the social behaviors; a second was comprised of the academic classroom behaviors and teacher ratings on IQ, mathematics, and English; and the third was comprised of the test scores in IQ, mathematics, and English together with the corresponding teacher ratings.

The question under investigation, the relationship of teacher ratings to standardized tests, goes back to Binet. However, opportunity for new illumination into this area was provided by a nationwide, longitudinal, societal experiment conducted in the Republic of Ireland to assess the impact of introducing standardized testing on various populations, particularly students and teachers (Airasian, Kellaghan, & Madaus, 1971). For the present study, 170 fifth-grade teachers were asked to provide ratings for each of their students on a variety of behaviors using a 5-point rating scale. Just prior to making their ratings, teachers administered standardized tests of intelligence,

Portions of this paper were presented at the annual conference of the National Council on Measurement in Education, Toronto, 1978. Work on this study was supported by funds from the Carnegie Corporation of New York, the Russell Sage Foundation, the Spencer Foundation, and the National Institute of Education. The views expressed are the sole responsibility of the authors.

mathematics, and English to their students. The tests (Educational Research Centre, 1973), which were built specifically for the Irish population of school-aged children and normed on a nationwide random sample of these children, were similar to counterpart U.S. tests. However, the newness of standardized testing in Ireland is a crucial aspect of the study. Teacher ratings were ensured of being free of influence from standardized test results, since teachers had not seen or administered such tests prior to the start of this study.

ANALYSIS

The test scores and ratings were factor analyzed using common factor analysis with iteration, squared multiple correlations for initial communality estimates, and a varimax rotation. An oblique rotation was also performed; the results were so similar to the varimax solution as not to warrant reporting them. Factors with eigenvalues greater than 1.0 were rotated. Only those cases for which there were complete (i.e., nonmissing) data for all the variables in the analysis were retained. This procedure for eliminating cases with missing data reduced the number of cases from a total of 4,362 cases with at least partial data down to 2,617 cases with complete data, a loss of 40.0 percent of the cases.

Examination of the configuration of missing data showed that for 10.3 percent of the total cases (448 students), teachers did not rate pupils' intelligence, but did rate all other areas, and test scores were available on those students. Although this attrition is quite high, the context of the study goes a long way towards explaining teachers' reluctance to rate intelligence. The teachers had little reluctance to rate English or participation in class, for example, since these behaviors were familiar to them. Intelligence, on the other hand, was a more abstract concept, one that they had probably never been asked to judge in any formal way before, and they were more unsure and reluctant to rate it. Missing test scores due to absenteeism were the other chief source of data reduction—with 20.1 percent of the total cases (or 878 students) being absent for one or more tests. No one test score was missing more than others. The remainder of the cases eliminated did not exhibit any particular pattern of missing data.

The 12 ratings of behaviors not directly measured by the tests were included in the factor analysis to provide an indication of both the degree of overlap of these variables among themselves and the degree of overlap among these variables, the teacher ratings, and the standardized test results. These 12 ratings have been subjected to extensive factor analysis by themselves, and have consistently revealed two distinct factors (Airasian, Kellaghan, & Madaus, 1977). Factor 1 relates to classroom behaviors and is comprised of items 7–13 in Table I. Factor 2 relates to social or personal student behaviors and contains items 14–18 in Table I.

TABLE I
Varimax Rotated Factor Matrix and Correlations Among Variables

Variable	Factor Matrix				Correlation Matrix																		
	Factor 1 Socia- bility	Factor 2 Aca- demic Class- room	Factor 3 Ability/ Attain- ment	<i>h</i> ²	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
<i>Tests</i>																							
1. I.Q.	19	22	<i>84</i>	79	79	79	75	61	64	62	55	53	53	50	55	42	45	29	27	31	32	30	
2. English	20	25	<i>78</i>	71		68	71	61	65	59	55	53	53	50	56	43	47	31	28	32	32	28	
3. Mathematics	19	21	<i>77</i>	68			64	58	58	63	52	50	51	46	49	41	42	28	26	32	32	27	
<i>Ratings</i>																							
4. IQ	13	<i>60</i>	<i>59</i>	72			78	86	81	68	64	63	59	64	54	58	33	30	38	36	30		
5. English	17	<i>60</i>	<i>61</i>	77				82	80	69	67	66	62	70	59	60	36	34	40	39	32		
6. Mathematics	17	<i>57</i>	<i>60</i>	71				74	68	65	63	60	61	55	60	35	33	37	37	32			
7. Attention span	43	<i>74</i>	<i>34</i>	84					84	89	84	74	74	70	82	56	50	53	51	39			
8. Persistence in school work	48	<i>74</i>	<i>30</i>	87						89	90	73	73	74	83	60	55	53	53	41			
9. Keeness to get on	48	<i>70</i>	<i>31</i>	82								84	73	75	71	80	58	55	54	54	41		
10. Participation in class	34	<i>63</i>	<i>34</i>	63									64	71	57	66	45	39	50	42	37		
11. Speech/use of language	40	<i>61</i>	<i>39</i>	69										71	67	69	46	47	54	52	38		
12. Neatness in school work	52	<i>56</i>	<i>25</i>	64											64	72	55	54	52	58	39		
13. Working with limited supervision	52	<i>67</i>	<i>23</i>	76												76	61	57	55	50	37		
14. Behavior in school	78	<i>24</i>	<i>11</i>	68													67	79	55	53	36		
15. Manners/politeness	88	<i>15</i>	<i>10</i>	80														70	64	55	39		
16. Getting along with other children	63	<i>29</i>	<i>17</i>	51																52	52	36	
17. Personal appearance	59	<i>29</i>	<i>18</i>	47																	47	41	
18. Attendance	40	<i>22</i>	<i>20</i>	25																		25	
Percent of total variance	21.9	25.7	21.1	68.6																			

Note. Decimal points are omitted. Squared multiple correlations are listed along the diagonal of the correlation matrix. Number in italics = factor loadings greater than .50.

RESULTS

Table I presents the factor pattern and correlation matrix. All factor loadings greater than .50 in Table I appear in italics. The two highest loading variables on Factor 1 are manners/politeness and behavior in school. The remaining high-loading items (getting along with other children, personal appearance and dress, neatness in school work, and working with limited supervision) relate to these two behaviors in a way that connotes a construct of sociability. This factor consistently emerged in the prior study (Airasian, Kellaghan, & Madaus, 1977), with the exceptions being that attendance loaded somewhat higher and neatness in school work and working with limited supervision somewhat lower in previous analyses.

The highest loading variables on Factor 2 are attention span, persistence, and keenness to get on (the latter being an Irish idiom for "desire to do well"). These variables relate to a dimension which can be named "academic classroom behaviors," and the remaining high-loading variables attest to the academic-related component of this factor.

Factor 3 contains the three test scores and the corresponding teacher ratings. This factor has thus been named ability/attainment. The three teacher academic ratings (IQ, English, and mathematics) load equally high on Factor 2 along with the classroom behavior items that replicate the results of previous factor analyses.

The new information obtained from the results of the analysis in this study rests in where the three test scores and the three corresponding teacher ratings load. The three test scores all load together on Factor 3 with the teacher ratings in these same three areas. Teacher ratings, however, load with the academic classroom behaviors (Factor 2) as highly as they do with the test scores (Factor 3).

DISCUSSION

Much of the teacher expectation/self-fulfilling prophecy literature assumes that artificial test results can change teachers' existing expectations for students. Little information exists on how actual test results relate to teachers' existing expectations, even though it is crucial to know this relationship in order to assess the potential for test results to affect expectations.

Teachers' existing expectations, as indicated by their ratings for their students' IQ and attainment, independent of any standardized test score results in these areas, tap a dimension similar to the tests. The introduction of one new piece of information, actual performance on standardized tests, would therefore tend to concur with and thus reinforce, on average, the existing expectations teachers held for students in these areas. Thus, on average, the potential impact of standardized test results on teachers' existing expectations appears to be minimized by the overlap between test results

and teacher ratings as indicated by Factor 3 in Table I. This result does not imply that the impact may not be great for an individual student.

The results also indicate that teacher judgments of students' IQ, English, and mathematics performance are confounded with their judgments of other academically related behaviors, such as attention span and persistence. Not surprisingly, teachers can not separate their judgments about academically related pupil behaviors which they observe on a daily basis from their judgments of pupils' standing on IQ, mathematics, and English. Standardized tests, on the other hand, are not confounded with these academically related behaviors to a great degree; they relate most strongly with teacher ratings in the same three areas.

Teacher ratings of social behaviors, such as manners/politeness and behavior in school, do not relate very highly with either the teacher ratings of IQ, mathematics, and English or the standardized test results in these three areas. Thus, social behaviors are not intertwined with teacher judgments of IQ, mathematics, and English to as great a degree as are academically related behaviors.

The major implication of the results from this study is that even in the absence of standardized test-score results, teacher judgments of students' intelligence and mathematics and English attainment tap a dimension similar to that tapped by standardized tests, but are also intertwined with academically related behaviors such as attention span and persistence. These other behaviors are ones that the teacher observes in students on a daily basis. It is not surprising that teachers cannot disentangle the purely cognitive behaviors from others that are related though not strictly of a cognitive nature.

REFERENCES

- AIRASIAN, P. W., KELLAGHAN, T., & MADAUS, G. F. *The consequences of introducing standardized testing: A societal experiment*. Dublin: Educational Research Centre, St. Patrick's College; Chestnut Hill, Mass.: School of Education, Boston College, 1971.
- AIRASIAN, P. W., KELLAGHAN, T., & MADAUS, G. F. The stability of teachers' perceptions of pupil characteristics. *Irish Journal of Education*, 1977, 11, 74-84. Educational Research Centre. *Drumcondra test series*. Dublin: St. Patrick's College, 1973.

AUTHORS

- JOSEPH J. PEDULLA, Senior Research Associate, Boston College, Chestnut Hill, MA. 02167. *Specialization: Testing and evaluation.*
- PETER W. AIRASIAN, Professor, Boston College, Chestnut Hill, MA. 02167. *Specialization: Testing, evaluation and policy studies.*
- GEORGE F. MADAUS, Professor, Boston College, Chestnut Hill, MA. 02167. *Specialization: Testing, evaluation, and policy studies.*