Are Teachers Really Poor Talent Detectors? Comments On Pegnato and Birch’s (1959) Study of the Effectiveness and Efficiency of Various Identification Techniques

François Gagné
Université du Québec à Montréal

Abstract

This article reanalyses Pegnato and Birch’s data in their often cited 1959 article, especially their measures of effectiveness and efficiency. The author shows that these two concepts are not independent, as usually assumed in studies examining the criterion-related validity of various identification techniques for the gifted and talented. Rather, both are related to the number of subjects singled out as a result of the cutoff value chosen to implement the method. Consequently, most comparisons made between and within methods become invalid.

A simple alternative is presented: compare the correlation coefficients of the predictors with the criterion. It is concluded that this methodological flaw, as well as some other technical limitations, invalidates the most often cited observation from that study, namely, that teacher nominations should not be relied upon as an identification technique for the gifted and talented.

The Pegnato and Birch (1959) study is among the most frequently cited in the literature on the identification of gifted students. In fact, this study appeared more than 25 times in the Citation Index covering the last 20 years. Gowan (1977) placed it among his choice of 12 research milestones in the field of giftedness for the 1950-1975 quarter-century, affirming that this study “laid the basis for sound research on identification” (p. 19). Moreover, the two concepts of effectiveness and efficiency adopted by Pegnato and Birch to organize their data analysis have been widely used in subsequent studies (e.g., Ashman & Vukelich, 1983; Gear, 1976; Jacobs, 1971; Waters & Clausen, 1983). Pegnato and Birch’s (1959) study is usually cited to buttress the opinion that teachers are not very reliable when asked to identify their more gifted pupils (e.g., Baldwin & Gargiulo, 1983; Clark, 1983; Freeman, 1985; Gear, 1976; Robinson, 1990). But do the data gathered by these researchers clearly show teachers to be poor judges of pupils’ abilities?

The authors’ objective was to compare “the relative efficiency and effectiveness of seven different means of locating gifted children in junior high schools” (Pegnato & Birch, 1959, p. 301). The criterion “gifted children” was operationalized as an IQ of 136 or more on the Stanford-Binet. The effectiveness of a method was operationalized as “the percentage of gifted children it locates”; its efficiency was the “ratio between the total number of children it refers for individual examination and the number of gifted children found among those referred” (Pegnato & Birch, 1959, p. 302). In order to obtain a sample containing a relatively high proportion of gifted children, Pegnato & Birch chose a large junior high school which was “situated in and drew upon a very favored group of neighborhoods from a socio economic standpoint” (1959, p. 301); the population was composed of the 1,400 students in Grades 7 to 9. Seven identification methods were used: (a) teacher judgment (nominations of their “mentally gifted” students); (b) honor roll (an all-subject average of B or higher); (c) creativity in arts or music (nominations by art and music teachers); (d) student council membership (those selected in their home rooms to represent their peers); (e) superiority in mathematics (nominations by arithmetic teachers); (f) group intelligence test results (an IQ of 115 or more on the Otis Quick-Scoring Mental Ability Test, Beta Form); (g) group achievement test results (average of the reading and arithmetic subtests of the Metropolitan Achievement Test, with a cutoff at least three grade levels above grade placement).

By pooling the lists, 781 different children were screened, that is, 56% of the population. All were administered the Stanford-Binet; 91 (6.5%) were found to have an IQ of 136 or more. From these data, indices of effectiveness and efficiency were calculated; they appear in Table 1. Here are most of the observations that Pegnato and Birch, (1959), drew from their analysis of the data: (a) “teachers do not locate gifted children

Putting the Research to Use

In the light of this reexamination of Pegnato and Birch’s (1959) study, teachers and program coordinators in gifted education should reassess their personal views concerning the accuracy of regular classroom teachers as “talent scouts.” Moreover, since many of these teachers are probably aware that specialists in gifted education have been judging rather negatively their abilities to identify gifted and talented children accurately, they should be informed that the main evidence which buttressed this negative judgment has been seriously questioned and that a closer look at the research of the last 20 years suggests much more positive judgments. Adopting and disseminating these more positive views of the usefulness of teacher nominations should help strengthen the ties between gifted education specialists and regular classroom teachers.
effectively or efficiently enough to place much reliance on them for screening" (p. 303); (h) the honor roll is very effective but “among the less efficient screening methods” (p. 303); (c) those few identified through art teachers and student council membership all appeared in at least two other lists; (d) math teachers were not good at screening gifted children ("almost half of them were overlooked," p. 303); (e) “group intelligence tests like the one used in this study cannot be relied upon in the identification of gifted children . . . [but ii] does seem to possess the best combination of efficiency and effectiveness as a screen” (pp. 303-304); (f) "group achievement test scores run a fairly close second . . . in combined efficiency and effectiveness” (p. 304); (g) "taken together, the two group tests resulted in the most effective screening procedure [88/91]” (p. 304).

Table 1
Effectiveness (Fv) and Efficiency (Fc) of the Seven Screening Methods Used in Pegnato and Birch’s (1959) Study

<table>
<thead>
<tr>
<th>Screening methods</th>
<th>No. nominated S B &gt; 135</th>
<th>No. of Phi</th>
<th>Fv %</th>
<th>Fc %</th>
<th>coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher judgment</td>
<td>154</td>
<td>41</td>
<td>45</td>
<td>27</td>
<td>.29</td>
</tr>
<tr>
<td>Honor roll</td>
<td>371</td>
<td>67</td>
<td>74</td>
<td>18</td>
<td>.28</td>
</tr>
<tr>
<td>Creativity</td>
<td>137</td>
<td>14</td>
<td>15</td>
<td>10</td>
<td>.05</td>
</tr>
<tr>
<td>(Art)</td>
<td>66</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>.02</td>
</tr>
<tr>
<td>(Music)</td>
<td>71</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>.04</td>
</tr>
<tr>
<td>Student council</td>
<td>82</td>
<td>13</td>
<td>14</td>
<td>16</td>
<td>.10</td>
</tr>
<tr>
<td>Math achievement</td>
<td>179</td>
<td>50</td>
<td>55</td>
<td>28</td>
<td>.33</td>
</tr>
<tr>
<td>Otis IQ ≥ 115</td>
<td>450</td>
<td>84</td>
<td>92</td>
<td>19</td>
<td>.34</td>
</tr>
<tr>
<td>(IQ ≥ 120)</td>
<td>240</td>
<td>65</td>
<td>71</td>
<td>27</td>
<td>.38</td>
</tr>
<tr>
<td>(IQ ≥ 125)</td>
<td>105</td>
<td>40</td>
<td>44</td>
<td>38</td>
<td>.37</td>
</tr>
<tr>
<td>(IQ ≥ 130)</td>
<td>36</td>
<td>20</td>
<td>22</td>
<td>56</td>
<td>.32</td>
</tr>
<tr>
<td>Achievement test</td>
<td>335</td>
<td>72</td>
<td>79</td>
<td>21</td>
<td>.34</td>
</tr>
</tbody>
</table>

Discussion

The indices of effectiveness and efficiency are implicitly treated by the authors as distinct and independent. However, Table 1 shows a high negative correlation between the Effectiveness (Fv) and Efficiency (Fc) columns; in fact, both concepts as operationalized are directly related to the number of subjects singled out by the method. This double association can easily be observed in Table 1 by comparing the effectiveness and efficiency levels of a given method (e.g., method X is very effective, but not very efficient) because these two indices will move in opposite directions as we change the cutoff score. This is most evident with the four Otis cutoff scores used by Pegnato and Birch; which of the four—quite different—pairs of indices should be used to assess the value of that predictor? Nor should we directly compare the effectiveness or efficiency levels of two methods unless we also take into account the number of pupils named by each. For instance, it is not fair to compare the effectiveness of the teachers to that of the honor roll because twice as many students were named by the latter; conversely, one cannot compare the efficiency of the teachers to that of the Otis ≥ 130 cutoff. Because four times fewer pupils were screened through that cutoff one should at least mention that teachers were much more efficient. We cannot even compare the amount of change in one index to that in the other because they do not change at the same rate. As shown in Table 1, effectiveness levels vary much more than efficiency levels.

Is it possible to neutralize the impact of the number of subjects singled out by a method? A very good and simple solution consists in computing a correlation coefficient between each method and the criterion. It will then be easy to assess the relative worth of each method by comparing these
correlation coefficients, even testing the significance of differences between them. Using Pegnato and Birch's (1959) data, it was possible to create a 2 x 2 table for each predictor against the criterion, then compute a phi coefficient, mathematically equivalent to a Pearson r applied to dichotomous variables (Ferguson, 1976). These phi coefficients appear in the rightmost column in Table 1. Using an α level of .05, only two of the Otis subgroups (≥ 120 and ≥ 125) had a statistically higher phi coefficient than the teacher nominations. Analyzed this way, the teachers do not come out worse than most other sources of information, including some subgroups of the Otis.

The conclusion of this short analysis is quite clear: educators in the field should stop citing Pegnato and Birch's (1959) study as proof of poor teacher judgment in identifying gifted and talented children; their data do not support such a sweeping judgment. Furthermore, a quite different picture appears to emerge from more recent research on this subject. As Hoge and Cudmore (1986) conclude from their detailed review of the literature on teacher-judgment measures, "there is very little empirical foundation for the negative evaluations so often associated with these measures. . . . In fact, the thrust of more recent research seems to be in a direction supportive of the judgments" (p. 191). The effectiveness and efficiency concepts were shown to give rise to major interpretational biases when operationalized as Pegnato and Birch did. Space permitting, we would have described other important measurement problems associated with these two concepts. But the limitations described above are serious enough to justify putting these concepts aside in favor of the more classical correlational approach.

Author's Note

The author sincerely thanks his colleague Helga Feider, who revised his efforts at writing directly in English.

References


