

Accountability That Counts

JOSEPH MILNER, CONNIE PULLUM COKER, CHRISTY BUCHANAN, DEBBIE NEWSOME, JONATHAN MILNER, RODNEY ALLEN, and MELISSA WILLIAMS

Abstract: The North Carolina Governor's School offers a six-week residential summer program for four hundred academically talented rising juniors. This article measures the school's impact on these students in four fundamental areas: cognitive maturity, moral reasoning, personal learning style, and projections for the future. The results showed that Governor's School students advanced further on each of these critical tests than did equivalent nonattending students.

Keywords: assessment, gifted education, residential summer program

These are the best of times and the worst of times for gifted students. Education leaders like Senator Lamar Alexander (2005) are particularly aware of the need to support and challenge gifted students so they can help the nation meet the rigors of global competition. Alexander has particularly supported programs that integrate content area knowledge across disciplines. Most governmental funding, however, is directed to programs like No Child Left Behind (NCLB) that primarily measure basic academic achievement of the least able students. Measuring such performance is important, but it can be a double-edged sword, especially in schools for gifted education. Educators must know whether they are achieving their goals, but mea-

suring academically talented students' performance is difficult. Narrow performance-based results should not constrain unique programs, however, just because of that difficulty.

North Carolina Governor's School, which opened in 1963, sets challenging goals for students and has gathered solid evidence of their success, earning the approval of the state legislature and the public. Each summer, four hundred academically gifted students from across the state are invited to attend the Governor's School for a six-week residential program. The school has developed an unusual curriculum that focuses on contemporary knowledge and theory in five core academic disciplines and five performing arts (Area I), the integration of knowledge across discrete fields of inquiry (Area II), and the exploration of the self in society (Area III; Milner and Milner 1986; Ward 1968). The curriculum focuses on the synthesis of ideas from diverse fields of knowledge and encourages the movement from fact to theory, that is, generalizing from the particular. Student evaluations have consistently offered high praise throughout the school's forty-plus years of existence. Over the past decade, 93 to 95 percent of a total four hundred students strongly agreed that "Governor's School was a rewarding experience" (Milner 2007). Few academic programs that operate during the summer receive anything close

Joseph Milner is a coordinator of the English Education Program at Wake Forest University, Winston-Salem, NC. Connie Pullum Coker is the technology support coordinator at Charlotte Latin School in Charlotte, NC. Christy Buchanan is a professor of psychology at Wake Forest University. Debbie Newsome is an associate professor in the Department of Counseling at Wake Forest University. Jonathan Milner is a member of the social studies faculty at North Carolina School of the Arts in Winston-Salem. Rodney Allen is the dean of students at North Carolina Governor's School, West, and an English teacher at Reynolds High School in Winston-Salem. Melissa Williams is a graduate assistant in the Department of Counseling at Wake Forest University.

Copyright © 2009 Heldref Publications

to that approval rating. Nor does time diminish students' regard for the program. Two years after completing Governor's School, the students still advocate vigorously for the program. After interviewing outstanding North Carolina students for the Morehead Scholars program, director Charles Lovelace reports, "We continue to be in awe of the power of the North Carolina Governor's School. Every year during our Morehead selection interviews, we hear young people from across the state assert that Governor's School was the most important and formative experience of their lives" (Lovelace, correspondence with author, May 21, 2000).

These accolades speak well of the Governor's School program, but because self-report can become self-congratulation, Governor's School supporters wanted a more objective assessment of the school's impact conducted. The desire to plumb the depths and measure the heights, rather than track minimal standards, led evaluators to probe four essential areas of focus at the Governor's School that closely related to the school's original mission. These four areas were central to the school's curriculum but could not be tested by high-stakes assessments that focus on narrow knowledge and ignore the deeper mental operations emphasized by the program. The four areas targeted cognitive growth, moral development, learning-style preference, and assessment of future expectations and global awareness.

Bedrock Measures: Four Fundamental Questions

This study compared Governor's School students to equivalent, nonattending students in terms of four fundamental questions.

1. Do students involved in this intensive curriculum, which offers extended seventy-five-minute classes that meet twice a day in a primary discipline or performance area, increase their cognitive maturity more significantly?

2. Do Governor's School students advance farther in their moral development because of the philosophical, epistemological, and axiological issues that are raised in their courses, integrated class sessions, and optional seminars?

3. Do Governor's School students more frequently move from a personal learning style characterized by concrete and sequential preferences to one characterized by abstract and global preferences?

4. Do Governor's School students more significantly advance from a narrow understanding of significant dimensions of life and important national and international issues to a broader understanding of life possibilities and the crucial national and international issues facing the world today?

Methods

Study Instruments

The research group began its assessment by considering cognitive growth, although changes in this area are difficult to track over a six-week program. We used Slosson's Arlin Test of Formal Reasoning (ATFR), which has been widely used in research on adolescents' cognitive development (Arlin 1982), but because we were assessing four growth areas, we limited our test items to a small part of the ATFR inventory. To measure moral development, we used the University of Minnesota's Defining Issues Test (DIT), which was used in an earlier study of moral development at the Governor's School (Milner, Ferran, and Martin 2003; Perry 1986). We used the Kaleidoscope Profile from Performance Learning Systems to measure learning-style preferences, such as Abstract/Global and Concrete/Sequential, through student self-report on a masked profile assessment. We constructed the final assessment, Future Inventory, to measure the students' life expectations, awareness of international and national issues, and self-assessment on a socio-philosophical continuum.

Design and Procedures

These four tests were administered to an experimental group of thirty-seven Governor's School students, who constituted 10.5 percent of the total population ($N = 389$). The experimental group was randomly selected from two Area III (exploration of the self in society) classes, because these classes included students from each of the Area I classes of the program. A control group was identified of fifty-five students who were accepted to Governor's School but declined to attend; only fourteen of the fifty-five completed the study, constituting about 25 percent of the potential control group.

A pretest consisting of four subtests was given to the experimental group at Governor's School during the first week of class, and the posttest was completed during the last week of class. The same six-week test schedule was used for the control group.

The four assessment measures were administered to the thirty-seven Governor's School students by a faculty member who had no information about the nature of the assessment or projected answers. The faculty member read the test instructions aloud, which clearly explained the process. The four assessments focused on different cognitive or psychological steps, and the whole set took thirty-five minutes or more to complete. Each test was coded for blind scoring; experimental and control group pre- and posttests were later separated to allow measurement of changes that had occurred over the six weeks.

Nature of the Four Subtests

Subtest 1: Cognitive maturity. The ATFR's foundations lie in developmental psychology. The test was developed

and published by Patricia Arlin at California State University at San Bernardino in 1982 as a measurement of cognitive maturity. The ATFR has demonstrated reliability and validity with large groups of students (Arlin 1982). We used only the portions of the instrument designed to measure students' reasoning capabilities and therefore selected two specific problems from the ATFR for our subtest.

The ATFR pretest focused on two dimensions of cognitive power using two word problems related to ratio and relationship, each with four possible responses. A correct answer earned students a point, with two points as the maximum possible score for the multiple-choice component. The explanation part of the test was an extension of the response students selected. Students were asked to explain their choices; one point was awarded for each explanation that was logical for a maximum possible score of four points on the ATFR. One of the two problems in this assessment is reproduced in appendix A.

The posttest used a similar format, with both problems worth one point and accompanied by written explanations. We anticipated that the scores of Governor's School students would show a greater increase from pretest to posttest than the scores of nonattending students.

Subtest 2: Moral development. The DIT is designed to measure moral judgment and is based on Lawrence Kohlberg's theory of moral development in which persons move from a preconventional, self-oriented stage of moral reasoning to a conventional, rule-governed stage to a postconventional, altruistic stage. The DIT is reported to have good psychometric properties and measures moral development less subjectively than Kohlberg's own test (Sternberg and Williams 2002; Sutton 2004). Concerns have been expressed that the test is dated (Sutton 2004), but it remains the most widely used test of moral judgment (Bebeau and Thoma 2003).

The DIT consists of six moral dilemmas and twelve related issues. For each of the six dilemmas, students select and rank the four issues they see as most important in resolving the dilemma from twelve choices. Appendix B illustrates a sample dilemma and set of questions. The posttest featured the same set of dilemmas. P scores, indicating the student's level of principled reasoning and represented by a percentage, were calculated from student choices of best responses. The test was digitally scored by the University of Minnesota's Center for Research in Ethical Development. The Humanitarian/Liberal subscore quantifies students' self-described stance on a 5-point continuum that ranges from 1 (*very liberal*) to 5 (*very conservative*). As with the ATFR, we anticipated a greater decrease in subscores among students enrolled in Governor's School.

Subtest 3: Learning-style preferences. The Kaleidoscope Profile measures learning-style preferences, rather than growth or change in knowledge or skills, and was developed by Performance Learning Systems, Inc. Further information can be found at <http://www.plsweb.com>. The Kaleidoscope Profile assesses an individual's learning style using three categories: sensory style, perceptual and organizational style, and personality style.

In this study, we focused on two opposing learning styles: Concrete/Sequential and Abstract/Global. Neither learning style was valued more by the instrument's developers. We chose these two opposing styles because of their relation to the Governor's School's primary objectives. We believed that significant movement toward the Abstract/Global style would indicate the fulfillment of the school's primary mission. The Kaleidoscope Profile also measures Sensory and Temperament styles, but although these areas help define students' styles, they were not relevant to our assessment. Their inclusion did, however, help mask the core preferences we were assessing, so students could not choose responses to please the test giver.

The process for this study was straightforward. Students were provided with four charts that each contained the beginning of a statement:

1. I enjoy school when . . .
2. School activities I enjoy are . . .
3. I value . . .
4. School is important because . . .

Students then selected up to eight stickers to complete each of the four statements and pasted them onto the charts. Examples of sticker choices for the first statement included *I can be creative; it is quiet; it is run well; and we don't waste time*. Responses to the second statement included *sharing opinions in class; taking notes; expressing my thoughts; and finding out about myself*. Responses to the third statement included *honesty; compassion; being right; and freedom*. Responses to the fourth statement included *it helps me learn and grow; my family believes it is; it develops skills for life; I hear about others' experiences; and of the opportunities*.

Each sticker was color-coded to indicate a learning-style preference, although students were not told the color-code system. Learning-style preferences as measured at the beginning of the six-week term were compared with preferences measured on the end-of-term posttest. We anticipated that Governor's School students would exhibit a greater shift from the Concrete/Sequential style to the Abstract/Global style than nonattending students.

Subtest 4: Future inventory. We developed the final assessment, the Future Inventory, for this study. The

inventory assesses student changes in three areas—ideas about the future, strong national and international concerns, and positions on a socio-philosophical continuum—over the six-week term. The first area deals with students' visions for their future using eight sub-areas, such as career possibilities, college and university choices, future residences, and possible family configurations. The second part of the inventory asks students to identify two national issues and two international issues about which they feel strongly. The final segment asks students to place themselves on a nine-point socio-philosophical continuum stretching from 1 (*very liberal*) to 9 (*very conservative*). This section nearly duplicates the DIT, which establishes a Humanitarian/Liberal mean score from students' self-selected responses to a philosophical continuum. The Future Inventory was completed at the beginning of the term and at the end of six weeks. We anticipated that the scores of Governor's School students would increase more on the posttest than the scores of nonattending students.

Results and Discussion

We tabulated pre- and posttest results for both the experimental and the control groups. Because of the small sample size of each group, we felt that inferential statistical analysis would be inappropriate; instead, we recorded mean scores for each of the four assessments to examine trends. These scores are shown in table 1.

As previously noted, the likelihood of students' cognitive powers changing markedly in six weeks was always low. However, differences between the experimental group and control group were evidenced by two important metrics. The experimental group's scores for the multiple-choice questions (out of a possible 2.000 points) of the ATFR rose 0.167 points, from 1.083 to 1.250. The control group's scores, although higher than those of the experimental group, increased by 0.070—less than half of the score increase of the experimental group. On the four-point scale, which included both the multiple-choice responses and the students' explanations of their answers, the experimental group's scores rose 0.083 points, from 2.167 to 2.250, whereas the control group's scores declined 0.140 points, falling from 2.640 to 2.500 on the posttest. These scores seem low, but had the students completed the entire ATFR, their projected scores would be around twenty-two, which would place them in the test's fourth highest group, "Low Formal: clear evidence of abstraction." Our results are commensurate with an earlier formal-reasoning test conducted at the Governor's School's 1985 summer session in which 30 to 35 percent of the students were classified in Piagetian terms as *formal operational* (Perry 1986).

The DIT results were easier to interpret. P scores improved from pretest to posttest for the experimental

TABLE 1. Mean Scores for Four Subtests

Test	Group	
	Experimental	Control
Cognitive maturity^a		
2-point results		
Pretest	1.083	1.430
Posttest	1.250	1.500
4-point results		
Pretest	2.167	2.640
Posttest	2.250	2.500
Moral development^b		
Pretest	43.17%	36.00%
Posttest	46.97%	36.49%
Humanitarian/liberal subscores		
Pretest	2.25	2.46
Posttest	2.25	1.79
Learning profile^c		
Concrete/sequential		
Pretest	5.88	5.20
Posttest	5.12	5.87
Abstract/global		
Pretest	7.32	6.73
Posttest	8.21	6.93
Future inventory^d		
Life projections		
Pretest	n/a	n/a
Posttest	2.000	1.285
Issues awareness		
Pretest	n/a	n/a
Posttest	0.945	0.642
Socio-philosophical continuum		
Pretest	4.303	5.071
Posttest	4.166	4.786

Note. Cognitive maturity tested using a modified form of the Arlin Test of Formal Reasoning (ATFR); moral development measured using a modified form of the University of Minnesota's Defining Issues Test (DIT); learning profile measured using a modified form of Performance Learning Systems' Kaleidoscope Profile; future inventory measured using a form created by the authors. ^aMeasured on a scale of 0–2 points, based on responses to multiple-choice questions. ^bEvaluated using P scores from the DIT. ^cMeasured on a scale of 1–10 from the Kaleidoscope Profile. ^dLife projections measured on a scale of 0–8; issues awareness measured on a scale of 0–4. Higher scores on the continuum indicate a more conservative outlook, whereas lower scores indicate a more liberal outlook.

group, whereas scores remained static for the control group. The experimental group showed an increase of 3.80 percent, from 43.17 to 46.97, but the control group's P scores rose only 0.49 percent, from 36.00 to 36.49. The score growth for the experimental group was more than six times greater than the control group's growth. The mean scores on the Humanitarian/Liberal continuum did not rise at all for the experimental group, but the control group score fell from a mean of 2.46 to 1.79. The experimental group thus outperformed the control group on both components of the DIT and displayed notable growth, whereas the control group's growth was minimal.

The results for the Kaleidoscope Profile were also clear. Eight points were possible on each of four questions, for a possible thirty-two points on the whole test. The experimental group's Abstract/Global scores rose by 0.89 points from 7.32 on the pretest to 8.21 on the posttest. The control group's scores increased only 0.20 points, from 6.73 to 6.93. Scores preferencing a less advanced learning style, Concrete/Sequential, fell for the experimental group, from 5.88 to 5.12, and rose for the control group, from 5.20 to 5.87, over the six-week term.

The final assessment, Future Inventory, produced similar results. Projections for the future and international and national concerns expressed by the experimental group suggested more expansive thinking about global issues and more awareness of national issues than the control group. The experimental group's self-placement on the socio-philosophical continuum also demonstrated a broad, open attitude. Compared to the control group, the experimental group showed a more pronounced shift from conventional choices to more unorthodox possibilities regarding future travel, graduate and professional study, future residences, careers, and family configurations.

Because of the open-ended nature of the Future Inventory, only changes (and not concrete responses) from the pretest to the posttest for each student in the group could be measured. If, for example, a student listed Chapel Hill and Asheville as possible residences on the pretest and then listed Chapel Hill and San Francisco on the posttest, that change would be considered significant and would merit a score of 1 for the section. Possible scores ranged from 0 to 8. The significant changes on the eight future markers of the inventory averaged 2.000 points for the experimental group and 1.285 points for the control group, a 64 percent difference. Possible scores for the issues section ranged from 0 to 4. Both groups showed growth on this section, but the experimental group's scores increased 0.945 points, whereas the control group's scores rose only 0.642 points. The socio-philosophical continuum provided a less obvious shift than the open-ended part of the inventory. Higher numbers indicate a more conservative stance and lower numbers a more liberal stance. The experimental group moved from 4.303 to 4.166, whereas the control group moved from 5.071 to 4.786. Both groups considered themselves somewhat conservative, but both moved about 2 percent closer to a liberal stance.

Importance of Findings

All four assessment areas of our study indicated that Governor's School students experienced more growth over the six-week term than equivalent nonattending students who spent the summer involved in other activities and events. Our findings can be summarized as follows:

1. On the ATFR assessment of cognitive maturity, the experimental group's gain score was 0.167 compared to the control group's gain score of 0.070. In the written part of the test, the experimental group's gain score was 0.083, whereas the control group registered a 0.140 loss. The experimental group thus outperformed the control group in both dimensions by more than 2 to 1.

2. On the DIT, the experimental group's P scores for moral development increased more than six times as much as the control group's scores over the six-week term.

3. The Kaleidoscope Profile of learning preferences indicated that the increase in the experimental group's Abstract/Global scores was four times greater than the score increase of the control group. Additionally, the experimental group's Concrete/Sequential score declined over six weeks, whereas the control group's score rose.

4. The experimental group outperformed the control group in both expansiveness of future projections and awareness of national and international issues.

It is a striking verification of the program's worth that in four deeply important ways, students at the North Carolina Governor's School outperformed equivalent nonattending students in intellectual growth and personal maturity. The original objectives of personal, conceptual, intellectual, aptitude, and integrative development lie at the center of these results. This assessment offers quantitative confirmation of the success of the Governor's School.

Limitations and Future Studies

Unlike logical or geometric propositions that can be proved, empirical studies can only accumulate evidence pointing to a possible truth. We do not suggest that the results can be generalized to a larger population, and we recognize certain limitations of this study. We could have included a larger percentage of Governor's School students, although 10 percent of the population is a reliable sample. The researchers also might have kept a closer connection with the fifty-five original decliners, so that all of them could have participated in the control group. Future studies should be conducted with other states' governor's schools that attract similarly bright students but have different curricula.

We believe the four measures of growth used in this study are important indicators of adolescents' academic and intellectual maturity, so we suggest that regular school systems adopt them to authentically assess student progress. Charting students' growth in these dimensions of their lives would balance the accumulation of data points that measure the more performance-based results that are so prevalent today.

REFERENCES

- Alexander, L. 2005. The National School Board Association's annual legislative conference. *C-SPAN*, January 31.
- Arlin, P. 1982. A multi-trait multi-method validity study of a test of formal reasoning. *Educational and Psychological Measurement* 43:103-9.
- Bebeau, M. J., and S. Thoma. 2003. *Guide for DIT-2*. Minneapolis, MN: Center for the Study of Ethical Development.
- Milner, J., J. Ferran, and K. Martin. 2003. Accounting for what counts. *The Clearing House* 77:62-66.
- Milner, J., and L. Milner. 1986. A conceptual approach to the humanities. *The Clearing House* 60:183-86.
- Milner, L. 2007. Governor's School West. Report submitted to the North Carolina Department of Public Instruction, Raleigh, NC.
- Perry, S. 1986. Moral Growth at North Carolina's Governor's School. MA diss., Wake Forest University.
- Sternberg, R., and W. Williams. 2002. *Educational psychology*. Boston: Allyn and Bacon.
- Sutton, R. E. 2004. Review of the Defining Issues Test. In *The eleventh mental measurements yearbook*, ed. C. W. Conoley and J. Impara, 268-69. Lincoln, NE: Buros Institute of Mental Measures.
- Ward, V. 1968. Differentiated education for the gifted. Paper presented at the annual meeting of the Council for Exceptional Children, New York City.

APPENDIX A

Sample Pretest and Posttest Questions—Cognitive Powers

A local toy manufacturer explained to one of his friends that no matter what he charges, the quantity he sells always goes up and down in relation to that price. The result is that the total revenue (money he takes in) is constant (remains the same). For example, if he raises his price, his sales decrease just enough so that the revenue is kept constant. Just the opposite happens if he lowers his price. His present cost per toy is constant no matter how many toys he produces. He wants to know how he should change the way he runs his toy business so that he can make the most profit.

Pretest Question: What can he do to make the most profit?

- A. There is nothing he can do because the total revenue remains constant.
- B. Pick a price that is right in the middle so demand is high but he can meet the demand.
- C. Increase the quantity of toys sold.
- D. Reduce the cost of producing the toys.

Explain your answer here: _____

Posttest Question: If the toy manufacturer were to double the number of toys that he made with no change in his cost per toy, what would happen to his profit?

- A. It would be cut in half.
- B. It would remain the same.
- C. It would double.
- D. It would be four times as great.

Explain your answer here: _____

APPENDIX B
Sample Test Questions—Moral Development

Mr. Grant has been elected to the School Board District 190 and was chosen to be Chairman. The district is bitterly divided over the closing of one of the high schools. One of the high schools has to be closed for financial reasons, but there is no agreement over which school to close. During his election to the School Board, Mr. Grant had proposed a series of "Open Meetings" in which members of the community could voice their opinions. He hoped that dialogue would make the community realize the necessity of closing one high school. Also he hoped that through open discussion, the difficulty of the decision would be appreciated, and that the community would ultimately support the School Board decision. The first Open Meeting was a disaster. Passionate speeches dominated the microphones and people threatened violence. The meeting barely closed without fist-fights. Later in the week, School Board members received threatening phone calls. Mr. Grant wonders if he ought to call off the next Open Meeting.

1. Is Mr. Grant required by law to have Open Meetings on major school board decisions?
2. Would Mr. Grant be breaking his election campaign promises to the community by discontinuing the Open Meetings?
3. Would the community be even angrier with Mr. Grant if he stopped the Open Meetings?
4. Would the change in plans prevent scientific assessment?
5. If the School Board is threatened, does the chairman have the legal authority to protect the Board by making decisions in closed meetings?
6. Would the community regard Mr. Grant as a coward if he stopped the Open Meetings?
7. Does Mr. Grant have another procedure in mind for ensuring that divergent views are heard?
8. Does Mr. Grant have the authority to expel troublemakers from the meetings or prevent them from making long speeches?
9. Are some people deliberately undermining the School Board process by playing some sort of power game?
10. What effect would stopping the discussion have on the community's ability to handle controversial issues in the future?
11. Is the trouble coming from only a few hotheads, and is the community in general really fair-minded and democratic?
12. What is the likelihood that a good decision could be made without open discussion from the community?

Copyright of Clearing House is the property of Heldref Publications and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.