Assessing High School Gifted Student Progress in Science

Through Misconceptions and MOSART

Christopher G Kolar

Evelyn Ho-Wisniewski

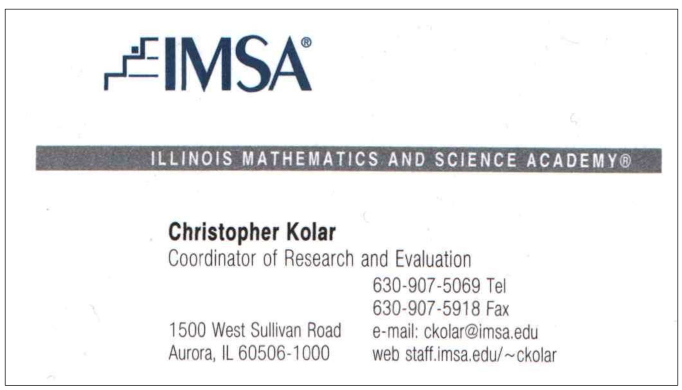
Illinois Mathematics and Science Academy

Presented at the 2009 American Educational Research Association Annual Meeting

San Diego, California

Abstract

This paper reports how 188 high school students identified as gifted in science were assessed with the Misconceptions-Oriented Standards-Based Assessment Resource for Teachers (MOSART). Students enrolled in a year-long science-centered curriculum where this instrument appeared to be a means of identifying standards-aligned progress, avoiding ceiling effects and reliance on content mastery. This paper discusses two questions: 1. Is the MOSART a valid measure of conceptual understanding in gifted students? and 2. Can the MOSART be used with this population to measure growth in understanding? We present results from the physics and chemistry tests, and consider results from the earth science and astronomy tests. We also discuss refinements to administration procedures and work expanding the subject pool in the coming year.



IMSA®

ILLINOIS MATHEMATICS AND SCIENCE ACADEMY®

Christopher Kolar
Coordinator of Research and Evaluation

630-907-5069 Tel
630-907-5918 Fax
1500 West Sullivan Road
Aurora, IL 60506-1000
e-mail: ckolar@imsa.edu
web staff.imsa.edu/~ckolar

Assessing High School Gifted Student Progress in Science Through Misconceptions and MOSART

This paper discusses a year-long project using the Misconceptions-Oriented Standards-Based Assessment Resource for Teachers (MOSART) instruments with a population of highly gifted high school students. Specialized programs for the gifted are often at a loss to find instruments that effectively assess growth in their populations. Ceiling effects from normed tests make it particularly difficult for such programs to demonstrate changes in achievement over time.

The MOSART instrument measures domain-specific conceptual understanding in either chemistry, physics, earth science, or astronomy. The tests are non-computational and are also aligned to the learning standards of their respective domains. These instruments are of particular interest because all distractor answers are based on domain misconceptions that have been reported in the literature. This means that incorrect responses allow for a possible diagnostic function by linking to the underlying research.

In this paper, 188 high school students attending tenth grade at a specialized secondary school for mathematics and science were administered one of the four MOSART instruments during their induction into the school, and then again at the end of their first year of instruction. Students apply to this school through a competitive process and attend from grades 10 through 12. Entering students have not only demonstrated academic achievement, but must also demonstrate ways in which they have pursued mathematics and science outside of school requirements. The class used in this paper entered the tenth grade with mean SAT scores of 652 and 587 for mathematics and critical reading, respectively (note that the assessment has been taken 3-4 years off-cycle), and have a prior GPA of 3.88 on a 4.0 scale. It is also important to note that students in this program do not participate in state-mandated testing. With students entering the program functioning at an extremely high level, the challenge was to find a valid means of assessing student growth in conceptual understanding of science.

The first year science curriculum is based on scientific inquiry across multiple disciplines. Students enroll in two courses each semester: inquiry-based courses in either chemistry, physics, or biology, as well as a course on methods in scientific inquiry. During the second semester students take the two courses not taken in the Fall. The curriculum is aligned to national standards in each of their respective disciplines, with the understanding that students will take on increasingly independent work as they become upper-classmen. For the purposes of this study, we used the MOSART instruments for high school physics and chemistry, as well as for earth science and astronomy. Because we did not know what to anticipate in student score distributions for this instrument or what changes in their scores would occur over time, the earth science and astronomy instruments served as controls for which we would expect no student growth as these subjects were absent from the curriculum during the treatment.

Instrument Viability

The first goal of this study was to determine if the distribution of scores on this instrument made it useful with a population of gifted students. To determine this, each of the 188 students was randomly assigned to one of four groups (stratified by gender) corresponding to the four subject tests. The tests were given to all students during induction testing prior to the beginning of instruction.

We found that all four MOSART instruments yielded results that were remarkably similar.

Table 1
Descriptive statistics for testing prior to instruction.

	Total Points Possible	Mean Score	Mean Percent Score	S.D.
Chemistry	22	10.64	48.37	4.05
Physics	25	12.81	51.26	4.08
Astronomy	16	8.23	51.42	2.95
Earth Science	20	12.45	62.26	2.61

Additionally, the scores for each of the subject tests had normal distributions. Of additional interest was student reaction, many approached us asking about test answers and whether they would be graded. Given that the students had near perfect prior academic records, it is unlikely that they had experienced such low expectancy for success in the past.

Measuring Growth

It was our intention that with this instrument, we would be able to measure changes in student conceptual understanding of science subjects, with the hope that it would be a valid measure of efficacy of an inquiry-based science curriculum. Our initial results using 188 matched pairs found no statistically significant difference between pre and post instruction administrations on the astronomy and earth science instruments, which acted as controls as these subject areas are absent from the first year curriculum. However, there were significant differences in the chemistry and physics tests, with scores increasing from pre- to post-testing.

Table 2
Matched pre- and post- pair results.

	Paired Samples t-test sig	Cohen's d	r
Chemistry	.002	.24	.175
Physics	.001	.28	.138
Astronomy	.536	.06	.032
Earth Science	.908	.01	.009

Discussion

Our initial finding that the instrument provides good score distributions encourages us to continue to look at how it may be used as part of our system of assessment. In the coming academic year, all entering students will take both the physics and chemistry tests. Earth science and astronomy will be dropped having served their purpose. It is our hope that the larger sample and more-sophisticated growth modeling will allow for more detailed examination of the relationship between student conceptual understanding and their performance in science courses.

Feedback from science faculty indicates that students at the end of the year may not have been putting forth as much effort. We feel that the uniformity of scores across administrations (r(Chem)=.739, r(Physics)=.839) show that the students were not engaging in random guessing or otherwise trying to sabotage their scores on a large scale. In the coming academic year, students will not take the post-test during testing-day in an auditorium, but will instead complete the instrument during the last week of the semester in their science class. Faculty feel that their presence and the classroom environment will spur performance.

Finally, informal work with the student test results revealed something unexpected in the distribution of answers. We found that across all four instruments, student responses appeared similar to the distribution of incorrect responses in the normative population. While the relatively small group sizes and the large number of questions would make statistical comparison unreasonable, in the coming year we will look closely at the distribution of responses in the expanded test-taking population and also look more closely at what conceptual stumbling blocks are revealed in their responses to see if they are aligned with features of the curriculum.

Author Note

The four Misconceptions Oriented Standards-Based Assessment Resource for Teachers (MOSART) instruments for secondary school students were at the end of their development cycle when we chose to explore their usefulness with this unique student population. The instruments were already used by many of the Department of Education MSP projects in Illinois, the authors of the instruments inform us that publication of studies of the instrument and its development are forthcoming.

We anticipate access in the near future to normative data on item response characteristics from the general populations that have participated in administrations of MOSART. We hope to make a more thorough examination of what appears to be an initial similarity in response patterns between our students of exceptionally high ability and those of the normative population.

As noted above, another wave of administrations has begun since initial submission of this paper. The first half of the second wave of students has completed pre- and post-administrations of both the physics and chemistry instruments. In the first half of this cohort we have again found significant difference between administrations with similarly small effect sizes (physics n=110, d=.13, chemistry n=117 d=.23). It is our plan to use the full second wave, completing the instrument in May 2009 and doubling our participants, in more-sophisticated modeling of student performance. Also, we have not seen any differences in potential spurious student responses now that the administration is being conducted in a classroom environment. In the future we will be able to include course performance data as well as student background characteristics and test scores in the development of more sophisticated growth models.

The use of these test scores as a measure of conceptual understanding is understood by the authors based on item construction derived from content-area standards explicitly addressing conceptual understanding of their respective areas. What has surprised us about the results to-date is that the subject area courses taken by the students were explicitly designed around the same learning standards used for MOSART construction. We hope that access to the response patterns in the normative population will help to inform us about student misconceptions in our target population and allow for the refinement of instruction to better restructure student conceptual understanding.

The authors would like to thank the developers of the MOSART instrument, Hal Coyle and Mary Dussault from the Harvard-Smithsonian Center for Astrophysics. Their assistance through consultation and access to pre-publication data on instrument development was invaluable. We would also like to thank our four anonymous reviewers for their valuable suggestions related to this early stage of our research.