View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Scholarworks@GVSU

Language Arts Journal of Michigan

Volume 12
Issue 1 *The Multicultural Classroom*

Article 12

1-1-1996

Assessing Writing: A Response

Ronald A. Sudol

Follow this and additional works at: http://scholarworks.gvsu.edu/lajm

Recommended Citation

Sudol, Ronald A. (1996) "Assessing Writing: A Response," *Language Arts Journal of Michigan*: Vol. 12: Iss. 1, Article 12.
Available at: http://dx.doi.org/10.9707/2168-149X.1508

This Article is brought to you for free and open access by ScholarWorks@GVSU. It has been accepted for inclusion in Language Arts Journal of Michigan by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Assessing Writing: A Response

Ronald A. Sudol

In the Fall 1995 issue of *Language Arts Journal of Michigan*, Faye Kuzma, Brenda Vasicek, and Lynn Chrenka offer a critique of some aspects of Michigan's High School Proficiency Test (HSPT) in writing, noting in particular the apparent slighting of such elements of the writing process as peer review and revising (98-104). As a member of the team that created the assessment framework, I would like to respond to several points raised in the article. At the outset, however, let me say that I entirely agree with the authors' ideas about the teaching and learning of writing. The pedagogy they espouse is completely sound and should certainly be considered for adoption by teachers interested in helping their students perform well on the writing assessment. But I am quite a bit less comfortable with the assumptions they make about what assessment is and what it can accomplish. The authors seem to expect the assessment instrument to not only measure the quality of the writing produced, but to structure the writing process of individual students, a task to which assessment is not well-suited.

When we talk about mandated, high-stakes, large-scale assessment, it is useful to keep attention focused on the way all of this is seen by the public, and by "public" I mean the entities that support public education through taxes and good will. We are asking eleventh graders (age about 16) to review some material on a selected topic, to do some off-the-top-of-the-head writing about it, to engage in some conversation, and then, later, to write an essay on that topic, taking up to 110 minutes to do so while being able to consult such standard reference books as dictionaries and handbooks. We expect them to produce a couple of pages of competent and polished writing—writing that makes and supports a point clearly and effectively in the judgment of trained readers.

Now, who is going to be brave enough to stand in front of the public and say that if students can't do well on such an assessment, it's because there's something wrong with the *test*, namely that the kids didn't get to exchange papers with peers, and the scheduling didn't *force* them to revise their work three times? The point is this: it's tough enough to reliably assess written *products*, but to try to assess *processes* would not only be nearly impossible but foolhardy as well. In designing an assessment such as the HSPT in writing, you have to follow a rule that is basic to many complex tasks: discover the boundaries of your job and don't cross them. If you stretch assessment beyond what can reasonably be assessed, the whole enterprise loses its focus. Although there's a connection between assessment and classroom techniques, these are essentially distinct activities. Let me expand on this idea that we ought to keep assessment confined within strict boundaries by responding to a few issues raised by Kuzma, Vasicek, and Chrenka.

Portfolios

Everybody loves portfolios these days. We on the management team certainly did when we began our task, and I am sure we all still do. At our first meeting in December, 1992, it seemed inevitable that Michigan would have a portfolio-type writing assessment for all the sound pedagogical reasons put forth by Kuzma et al. But assessment must answer to other masters besides pedagogy. Portfolios enjoy very high validity as a means of assessing writing. But the high validity comes with a price—lower scoring reliability. As validity rises, reliability falls, and vice versa. The opposite is true of the multiple-choice test: its validity for assessing writing is very low, but its reliability is very high because it's easy to agree on what the right answers ought to be. We rejected the multiple-choice test because of its low validity, and we ultimately rejected the portfolio because of its low reliability.

. . . assessment must answer to other masters besides pedagogy.

Low reliability means that scoring the portfolios in the context of a massive assessment project would be highly problematic. Imagine the scene: 110,000 portfolios (that's about the number of students in each grade) arrive in Lansing; each contains 10 to 20 pages or perhaps even more (and if you could stack them up, they would rise higher than the Sears Tower in Chicago); assuming each portfolio has to be read twice, how many readers would you have to hire, and pay, and train, and feed, and transport in order to read and score a total of between two million and four million pages of student writing? Even if you could solve the logistical problems and had unlimited funding, how would you train thousands of readers to be fair and consistent in scoring portfolios whose content is infinitely variable? Do you average the individual items? Does a ten-page research report have the same value as a one-page book report? Do you try to draw conclusions from the chronological sequence of the exhibits? It may be possible to solve these and countless other problems, but the solution hasn't been invented yet.

The fact is, portfolios are great in the classroom. They can be made to work building-wide, maybe district-wide and for college admissions and other specialized uses. But the further from the student and teacher you take the process, the more problematic it becomes. In 1990, anticipating a writing MEAP at some future date, the Michigan Department of Education sponsored a trial assessment that included several hundred portfolios from various grade levels in several school districts. After a long weekend of scoring, those of us who participated in this project discovered that while the portfolios were interesting to read, we could not agree with each other on scores. Indeed, the discrepancies were so great that we found ourselves using an impromptu piece of writing included in each portfolio to adjudicate discrepancies. It was a sobering experience.

The State of Vermont had a similar experience. Its proposed writing assessment involved an elegant use of portfolios. A study by the Rand Corporation of a pilot version of the assessment found inter-reader reliability to be unacceptable. If reasonably consistent scoring could not be achieved in a small and relatively uncomplicated state like Vermont (where the largest city is not even as big as Muskegon), imagine what it might be in the sprawling diversity of Michigan.

Another impediment is that the further away from the classroom you take the portfolio, the more likely you are to encounter questions of equity. All sorts of more or less legitimate complaints are possible: the portfolio disadvantages students who have not been in the same school during the previous year or two; teachers and schools might exert undue influence on the contents of portfolios since they have an interest in the outcomes; true authorship may, in some cases, be called into question; thus, students who can get the best help with their portfolios may be unfairly advantaged. I would like to think it's possible to overcome these obstacles at some time in the future. But for the present, it would be lunacy (and an invitation to litigation) to attempt to score portfolios until we have invented and perfected a reliable method for doing so. The

assessment as presently designed has a life span of three years. By calling for two pieces of writing to serve as the admission ticket to the exam, we have opened a door to the use of portfolios for this assessment at some future time, and we have encouraged the use of portfolios locally.

Peer Response

Like portfolios, the use of peer response and collaborative learning resonates very well among language arts professionals. Kuzma et al. are no exception. They extol the value of collaboration in facilitating the writing process and seem stunned by the paradox of a writing assessment that does not include structured time for peer review. Those of us on the management team had the same feeling when we began. Indeed, the first version of the assessment did include structured peer review. This part of the plan was widely criticized by Michigan teachers when we presented it at various field sites throughout the state in February 1993. The teachers attending these sessions were sensitive to the values of collaborative learning but highly skeptical about its fairness and viability in a high-stakes assessment. After much agonized discussion and deliberation, we had to agree with that consensus.

The ability to maintain this high standard is one way to avoid having any part of the test consist of multiple-choice editing questions.

Let's assume every positive claim about collaborative learning were true (highly dubious). In such a case, a certain percentage of the end product may be attributable to the peer collaborator. In school—as in life—that's the way it is. We do indeed make our way through school and through life collaboratively. It is certainly possible to assess the team or group effort and give the same score to every member of the team or group. But in the case of the HSPT we are concerned with an endorsement on a diploma, which is viewed legally as a personal possession, a property right representing personal achievement. When people win an Oscar or an Emmy, they thank all the folks who helped them along the way, but in the end the thing belongs to *them* and represents the recognized personal accomplishment of a single individual. The endorsed high school diploma may be easier to get than an Oscar, but like other awards and recognitions it's got only one name on it. Kuzma et al. correctly point out that the writing MEAP tests, unlike the HSPT, do include structured peer review (on the second day only). The reason is that the MEAP is not connected to a diploma. It reports a student's progress in a particular set of skills as part of a long process leading to the proficiency exam.

Apart from the legal status of a diploma, there are quite a few practical objections to structured peer review. Can we assume that every single student in the whole state will be ready for peer review at the same moment? If some students are having collaborative conversation, will they distract those who may still be drafting? Won't the best peer reviewers be in great demand and thus be unable to concentrate on their own writing? What about a student whose peer reviewer has marginal skills? Isn't that student disadvantaged? Isn't the socioeconomic composition of some classrooms an advantage or disadvantage compared to others? If a student fails, to what extent is the failure attributable to the collaborator? The three-day structure of the MEAP writing assessments and the flexibility of individual classroom settings make it possible to avoid or compensate for these constraints. But in the higher-stakes proficiency exam, these considerations could open a door to needless and distracting complaint and litigation.

Finally, though collaborative learning is clearly important, like so many excellent pedagogies, it tends to get fossilized in application. If, heaven forbid, the HSPT in writing included a lock-step structure like 10 minutes for brainstorming, 10 minutes for drafting, 10 minutes for peer review, 10 minutes for revising, and 10 minutes for proofreading—or any other structured device—every student in every classroom for every assignment is going to be put through the same regimen,

like it or not. Obviously, students work in a variety of ways, at different paces, with different needs at different times. Our plan allows them to structure their own time according to whatever process they have found works for them. An important role for the teacher is to help them discover what sort of process will work best.

Revising

The distinctive feature of the writing assessment is the 110-minute "extended" piece, written to a prompt based on (but different from) the focused reading, writing, and discussion of a particular theme or topic on a previous day. We had decided early in our deliberations that we did not want the assessment to depend on impromptu writing, where students are called upon to respond to a prompt in a limited time—discouraging revision and requiring the essay to be scored as a first draft rather than a finished piece of writing. The portfolio lacks these disadvantages, but for the reasons already mentioned, the portfolio was insufficiently reliable. The extended piece is the obvious compromise—it's based on a topic or theme that the student has already done some work on and has had some time to think about; it allows ample time for revising and rewriting; and it allows students to polish their work using dictionaries and handbooks so that it can reasonably be scored as finished writing. Since students have time for invention, drafting, revising, and polishing, readers need not give too many benefits of the doubt because of under development or error. The ability to maintain this high standard is one way to avoid having any part of the test consist of multiple-choice editing questions.

The extended piece is clearly a compromise between the very short impromptu and the very long portfolio, so it is no surprise that we catch flak from opposite poles. Some teachers tell us 110 minutes is way too long. The students write in about 40 minutes, spend 10 minutes making corrections, leaving a full hour for mischief-making. (One can only hope that the administration of the exams will permit students who have definitely finished their writing to do some other productive work.) The fact is, many students can perform very well in half the time allowed. The extra time is for those whose processing takes longer, who truly benefit from multiple drafts, and who may not discover where they are going until they get there and need to make extensive changes.

Kuzma et al. argue the opposite extreme—that even more time is needed, several days, presumably. Perhaps they too easily equate "engagement" with the amount of time students are forced to work on their writing. I think students will stay engaged only long enough to satisfy the demands of the task, which is quite properly defined in terms of the quality of the writing produced rather than the amount of time or number of days devoted to its completion. There is no obvious formula for determining how much time is enough for the specialized purpose of a state-wide assessment, but we felt a double class period was about the maximum we could expect of our students. Other non-standardized writing projects—and writing across the curriculum—surely deserve more time. The nine hours allowed for the assessments in all subjects is more than adequate to get the specific job done. It's hard to see how adding even more time could improve the accuracy and influence of the assessments.

We cannot depend on the assessment plan to guide the curriculum in any but the most general way—primarily by defining what is valued in written products.

Assessment depends on standardization—a hard fact often overlooked by people devoted to curriculum development. For example, Kuzma et al. ask "Why couldn't the HSPT . . . require students to write sustained, organic pieces of writing over a more extended period of time on a particular topic, one of their own choosing or one from a variety of listed topics. . . ?" The reason this can't be done is that in order to be reliably scored, every piece of writing must address the same task. Every variation from a fixed standard opens

the possibility of inequitable scoring. If we are going to avoid machine-scored writing tests, we need to be highly vigilant about the variables that influence human judgment. Standardization makes it possible to train raters to be fair and consistent in their scoring. Fail to standardize, and you invite litigation. The problem we face is to standardize in such a way as not to subvert the curriculum. But a standardized test cannot be a simple reflection of a good curriculum.

Conclusion

We cannot depend on the assessment plan to guide the curriculum in any but the most general way—primarily by defining what is valued in written products. The assessment framework studiously avoids conveying any notion that this is a test to teach to. *Any* sound teaching method or innovation should help students confront the rhetorical situations presented by the exam. I have read and scored literally thousands of student essays, at all three grade levels, written during the tryout and pilot stages of test development. I found the writing better than expected. There is obviously a lot of good writing instruction going on. When the writing is less than satisfactory, it has more often than not reflected a student's lack of motivation and skill in using the time allowed to good effect. Training and practice on how to use the generous time allowance is what students need most urgently. Their past experience with writing has often programmed them to believe that every additional sentence they write only provides another opportunity to make a mistake and lose points. They may not have had experience writing for holistic scoring, where they get credit for what they have done well. At the same time, it would be most beneficial for teachers to work in holistic scoring sessions in order to understand the scoring process and to form a consensus on what to value in student writing—a most potent form of professional development. The best news is that the assessment framework and the subsequent test development process has generated much useful conversation, and that should continue.