"My words fly up, my thoughts remain below" The challenge of assessing student learning in the Shakespeare classroom Barry Oreck Dawn McAndrews

How does acting in a Shakespeare play differ from reading it? What theatrical approaches will be effective in reaching students who have never read a Shakespeare play before? For theater companies working in schools, the answers to these questions are critical to understanding and demonstrating the value of their work as part of the language arts, classics, and drama curricula. Over the past three years, the Shakespeare Theatre Company has begun to study those basic questions through the development of a new student assessment process to better understand what students are learning through its Text Alive! program.

Text Alive! has been a mainstay of Shakespeare Theatre Company's education and outreach initiatives for more than 17 years. Working in 30 Washington D.C-area public high school classrooms each year, the program has a well-established curriculum, clear educational goals and objectives, and strong relationships with schools, most of which apply to participate in the program year after year. Feedback from participating English and Drama teachers, school administrators, students, parents, and the program funders has been consistently excellent; it is clearly a successful, valued program. So it was surprising to us -- and, we believe, instructive to others in the field -- to recognize the difficulty in answering one simple question: *"How do we know what individual students are learning?*

Wasn't it obvious? Through the program students have the chance to explore the language and history of Shakespeare and perform on a professional stage in front of their families and peers. For many it is their first chance not only to read but also to perform in a Shakespeare play, to learn and deliver lines from a classic text with understanding and feeling, to work as an ensemble. Students analyze Shakespeare's verse from a given play and as a class create a concept for their scene that makes sense for the characters, relationships, and setting. They create designs and costumes -- and select music to enhance their scene. Finally, the students, teachers, family and friends come to the Company's mainstage in downtown Washington where all participating students perform their scenes, resulting in an eclectic but united presentation of the play. Yet the question remained: *How, in the collaborative art form of theatre, could we infer or assess what individual students had learned and how much they had progressed?*

The process of trying to answer that question challenged the very core and structure of Text Alive! It forced teaching artists to confront their assumptions and look at their pedagogy. It changed the roles played by the collaborating teachers, and ultimately, we believe, resulted in a stronger program that is more responsive to students' needs and offers more opportunities for each individual to learn and demonstrate understanding. The goal: to create a systematic student assessment process that could capture student progress over time, would be practical to administer, and would maintain a focus on learning and engagement. In this article we trace the development of that process, which began in the Spring of 2004, and report on the results, even as we continue to improve and refine it, in the hopes that we can further dialogue in the field about methods and approaches for assessing students in a theatre arts residency program.

Systematic, individual student-level (as opposed to group) assessment poses many special challenges and raises serious questions for school residency programs sponsored by outside arts organizations. Time is short; Text Alive! consists of ten 45-90 minute classes (depending on the school) conducted over 10-12 weeks each semester. Teaching artists work with 2 to 6 classes of between 15 and 30 students. Classes vary greatly across schools. Knowledge and effectiveness of classroom teachers, grade levels, students' reading abilities and prior exposure to Shakespeare are just a few of the variables. So the challenge and the question was whether an assessment process

could be created that was adaptable to a wide range of school settings, that could challenge all students regardless of prior experience with Shakespeare or prior theater instruction, and that would not be so time consuming that it would undermine the program's instructional goals.

Our goal was not simply to test students' acquired knowledge of Shakespeare's text or a particular play. Text Alive! is based on the premise that dramatic involvement enhances students' understanding and appreciation of Shakespeare's work in ways that simply reading or identifying themes of the play do not. Students work intensively on a single scene from a single play. The goal: that they can then apply methods of analysis and interpretation to full plays and other texts as well as develop vocal and physical presentation skills. To evaluate that premise we could not simply give the students a written test. Students would need to demonstrate their ability to apply their newly acquired or enhanced knowledge and skills with new material in a dramatic context. Further, they would need the opportunity to express personal responses to the text in open-ended ways and to consider other aspects of theatre including sets, costumes, sound and stage design. A second question thus arose and became integral to the assessment: *in what ways does studying, designing, and performing in a Shakespeare play enhance understanding, knowledge and interest in Shakespeare*?

Methods

Program Design. Each semester Text Alive! involves 15 classrooms in 10-12 class sessions once a week for a 45-90 minute class period. A primary teaching artist, assisted by a Shakespeare Theatre Company education intern, facilitates the workshop at each school. Each semester focuses on a single Shakespeare play that the Company is scheduled to perform near the end of the semester. The curriculum is designed in two parts -- humanities and performance. In the first workshop, students are encouraged to get "on their feet". Vocal and physical warm-ups and ensemble building exercises get students comfortable moving and speaking with each other and prepare for an active examination of Shakespeare and his world. Through improvisations and discussion students explore the events of Shakespeare's life and the world of Elizabethan England. In sessions two through four students explore the language, history and imagery of Shakespeare's plays. Exercises focus on themes, poetry, and symbolism in the text while introducing students to the physical and vocal requirements for performing his plays. During this period students are reading the entire play in their language arts or drama class. Before the fourth session each class is assigned the scene they will perform as part of the final performance. Students work in design teams to conceptualize themes, characters, plot, and concepts for the production. They brainstorm a location and time period in which to set their scene to effectively communicate the playwright's intentions to a modern audience. This format allows all students to express opinions about the production and hear different artistic viewpoints. From the brainstormed concepts, each class ultimately chooses one for their scene that will support the action, text, and characters of the play. The group concepts are submitted as a mid-semester assignment that is assessed to gauge the students' understandings of the themes, characters, and action of the scene. The final four class sessions are primarily devoted to developing acting and vocal techniques, understanding each line and stage direction, rehearsing the scene for performance, and addressing any special requirements of the scene (such as stage combat, working with props and costumes, music, etc.).

Students come downtown to the theatre for a dress rehearsal onstage before the performance. On the morning of the student performance, all participants gather for a group vocal and physical warm-up and then each class performs their scene to create the play. Students return the following week to attend a performance of the professional production. After the performance the teaching artists return to the classroom to conduct the final assessment and to share thoughts

and perceptions about the student and professional productions.

Schools. The makeup of schools in the program changes year to year but approximately two thirds of the 12-16 schools participating in a given year are Washington DC public schools (a mixture of traditional and small charter schools) and one third are from suburban Maryland and Virginia. Schools participate in Text Alive! for an average of 3-5 years.

Students and Classes. Classes are selected for involvement in the program based on the interest of their teachers and schools. Teachers commit to four professional development sessions per semester and are responsible for ensuring that students read the play, and complete the journal and homework assignments. Most of the classes in the D.C schools are 10th, 11th and 12th grade English classes while most of the suburban schools tend to be Drama classes with a mix of 10th, 11th and 12th graders.

Assessment Process. To assess student progress over the course of the program a pre-post design was adopted involving performance-based observational assessments during sessions two and eleven. In addition, a text analysis assignment and a small group design project, journal prompts, written homework, and student surveys are collected at various times during the semester to track changes in students' attitudes toward Shakespeare, their confidence in approaching his work, and their engagement in their own learning.

Assessment Criteria. The teaching artists and education staff defined the assessment criteria to reflect a balance in the program goals and curriculum among three major elements: text understanding (content understanding and imagination), cooperative behaviors (focus and collaboration, related to their work in the theatre activities), and theatre skills (vocal and physical). A 5-point scoring rubric for the six items (two items for each of the three categories) was created and the descriptors were revised throughout the pilot process (Appendix A). Given this scoring distribution, the faculty felt that any student could excel, regardless of their innate acting ability or prior experience with Shakespeare.

Observational Performance Assessment. Initially, three observational assessments were designed. The first, conducted during session two, serves as a pre-assessment. After trying to conduct the pre-assessment in the first session, we concluded that it provided more accurate and valid information in session two when students showed greater comfort in the class. The final assessment, following the performance, demonstrates the students' understandings of the play as a whole, as well as the development of their analytical skills. There are three scorers– the school programs manager, the teaching assistant (usually a Shakespeare Theatre Company intern) and the classroom teacher; this allows the teaching artist to focus solely on leading the workshop and ensures reliability across all classrooms. The second assessment, conducted during session five, was dropped after two trials, as it seemed unnecessary given the other data collected throughout the program and the already limited instructional time.

Baseline data is collected during the second workshop in which each student is given a description of a person from Shakespeare's time along with a line of text to deliver with the physical and vocal characteristics suggested by the person's status, occupation, and attitude. After participating in a series of group improvisations, each student is given approximately five minutes to prepare the line and perform it twice with a brief comment or suggestion from the instructor following the first performance.

The second performance assessment, conducted during the final class visit, is similar to the first. After a group warm-up and activity, each student is given a line of text from somewhere in the play studied. Students are asked to deliver the line in character in such a way that the rest of the

class can identify the character, meaning, and moment from the play. Again the student receives a brief piece of feedback from the instructor following the first attempt. In this way, we are able to assess both the performer's and audience's knowledge of the play as they guess who, what and where based on the performance.

Applied Learning Project and Written Responses. In addition to the two performance assessments we wanted to look at other aspects of students' theatre skills and knowledge and to see how well they could apply the text analysis methods taught during the first four workshops.

In the Applied Learning Project, students work in small groups to create their own concept and overall design for the class scene. They can focus on costumes, lights, set, music, or overall design, presented in drawings, photography, models, writing, dramatization or other media and supported by a written and verbal explanation of the purpose and rationale for the choices made. This project allows us to recognize theatre skills other than acting, and to gauge students' levels of interest and engagement in the project at that point.

After the fourth workshop, students are given a scansion and paraphrasing assignment that they self-assess based on a four point rubric. This assignment gives the teaching artists and teachers immediate feedback on the students' grasp of the text analysis tools taught before moving into the more performance-focused phase of the program. In addition, throughout the semester, specific journal prompts and assignments are given to students and collected by the teacher.

Student Surveys. At the beginning and again at the end of the program students complete a survey concerning their attitudes toward Shakespeare, self-efficacy in reading and understanding the text, and overall feelings about their experience in Text Alive! The goal of the survey is to gauge changes in attitudes and efficacy over the course of the semester, to encourage students to think about their learning, and to gather open-ended responses from students about their Text

Alive! experience. The survey was edited over time based on student responses and includes both negative and positive statements to encourage thoughtful and honest responses. For example, students must rate their response (5 point scale) to both, "I usually understand Shakespeare's writing" and "I tend to be bored reading Shakespeare." Open-ended questions include: "Have you had any previous experience with Shakespeare's plays? Was it positive or negative? Why?" (pre-) and "What was your overall experience of Text Alive? What was your favorite experience? What was your least favorite experience? Please be specific." (post-). The post survey also includes another scansion exercise to see if the students can apply their skills to a new piece of text. Responses are tallied, pre-post- comparisons made for numerical items, and open-ended responses added to the student database.

Research Questions. The focus for the initial phase of research was to develop and test the assessment process itself – to prove its practicality and investigate its validity and reliability. We also had a number of basic questions about the variations among different students, schools, teachers, and teaching artists and the effect of the experience on student learning in the various settings. Further, before comparing the Text Alive! approach to other ways of studying Shakespeare it was necessary to identify the variables that could be expected to show the most change through the drama experience. The general questions about the process included:

- 1) Is the assessment process practical within the time frame of the classes?
- 2) Do the assessments themselves have educational value does learning take place during the assessment process?
- 3) Does the assessment reveal what the students actually learned in the program (as opposed to trivial or irrelevant variables)?

In addition, specific research questions included:

- 4) Do students show progress over the course of the program measured by the performance assessments? In what categories do students show the greatest and least improvement?
- 5) Are there significant differences in performance among schools in the program?
- 6) Are the assessments able to differentiate between levels of student performance? Is there a range or do scores tend to cluster together?
- 7) What is the level of agreement among the observers (interrater reliability)? Do the teaching artists and assistants tend to agree with each other and with the classroom teacher?
- 8) Are the results of the performance assessments supported by the other components of the assessment? Is the level of understanding and motivation demonstrated in the class activities reflected in the students' journal prompts, attitude surveys and design projects?

We analyzed the first set of questions about the process through ongoing discussions with the teaching artists, education staff, and participating teachers. Correlational and descriptive statistics were generated for the second set of questions. Statistical analyses included repeated measures ANOVA to ascertain the significance of the differences between pre and post scores (SPSS 11.0) and interrater reliability analysis to investigate the level of agreement between the teaching artists and among the teaching artists and classroom teachers. Writing and survey responses were read and organized on spreadsheets along with the observational assessment scores. Additional analysis was done for students who showed large increases or decreases in their scores to see if those results were consistent with the qualitative components of the assessment.

Results and Discussion

The assessment process has now been conducted six times and is fully integrated into Text Alive! Some, mostly minor, changes and adaptations have been made over the six administrations so some variation in the results can be attributed to changes in the process itself. Across all of the administrations, however, a number of strong trends have emerged that allow us to begin to draw conclusions about the process and form the basis for future research.

Observational Assessments

Overall, the assessment process was deemed very successful. Teaching artists and classroom teachers felt that the assessments were useful, revealing, educational experiences, and even enjoyable for the students. The marking procedure, somewhat daunting at first, became easier with practice and with some fine-tuning of the instructions and set-up of the assessment lessons. Time was the biggest obstacle, especially in the urban schools with 45-minute periods and large classes. In some of the suburban schools with longer or double periods and smaller class size, the entire assessment process could be conducted in a more relaxed way. The role of the school programs manager and education intern increased over time; the interns are able to watch the entire class without teaching responsibilities and can see the assessment process in multiple schools, giving them a broader perspective from which to rate student performance.

The statistical results of the pre-post performance assessments across all of the schools consistently showed improvement in student performance. Statistically significant improvement was seen in both urban and suburban schools with 65-75% of students increasing the scores across all categories and 75-85% showing improvement in at least 4 of the 6 criteria. Scores were higher overall (pre- and post-) for the suburban students who were primarily in theatre classes, as opposed to the urban students in significantly larger language arts classes. Higher initial scores may also have limited the suburban improvement slightly, although this has varied semester to

semester.

As would be expected, results have varied somewhat between different administrations of the assessment. Different plays, different schools, classrooms, teachers, and, in some cases, teaching artists, were involved each semester. The category that consistently showed the most improvement across administrations was Vocal Skills (projection and communication), one of the two theatre skills categories. The categories of Interpretation (generating and offering ideas) and Insight (expression of deeper understanding through personal choices) also showed significant improvement across all administrations and schools.

The difference between schools and administrations highlights the difficulty in interpreting statistical results in as multi-faceted an experience as theatre. When a student expertly acts his or her line with conviction, all of the criteria may be apparent at once. Interpretation, insight, focus, vocal and physical skills combine in the portrayal. It became clear in analyzing the results that improvement in specific categories was less telling than the overall understandings and ability to communicate those understandings through the voice and body. As such, the category of Vocal Skills may stand out as a proxy for a broader category encompassing confidence, comfort with the material, and desire to communicate with the audience. The separate categories remain useful, however, in providing clear language to focus the raters' attention and help distinguish between students, but the interrelatedness of theatre performance works against fine distinctions between the different criteria, particularly in examples of outstanding performance.

Investigation of Question Two showed that the assessments could differentiate between students at all performance levels. Total overall scores for the initial and final assessment fell into a normal curve (Tukey's estimation formula) and the three subcategories also fit into a normal distribution with Content Understanding and Behavior skewed somewhat toward the upper ends of the scale.

Multiple regression analysis was used to investigate which of the criteria made the largest unique contribution to the final overall score. Results varied somewhat between administrations but in all cases all six criteria contributed significantly to the final results and the most revealing criteria -- the one that most effectively evaluated differences between students -- was Vocal Skill. Focus was second and Insight or Interpretation (depending on the administration) third.

Analysis of Interrater Reliability (Question 4), concerning the level of agreement among the teaching artist, assistant, and the classroom teacher, showed a low to moderate level of agreement between the assessors (.5-.6). This was not unexpected as this statistic is strongly affected by the numbers of subjects – in this study the average number of students for each group of raters was less than 20, too few to expect statistical significance. However the result may point out the need for additional training for assessors and continued clarification of the criteria.

Validity Evidence

Content validity evidence was gathered through an extended development process involving a range of Shakespeare Theatre Company teaching artists and education staff and the partnering English and Drama teachers. This process involved both the creation of assessment criteria and behavioral descriptors and the redefinition and editing of the curricular goals and learning objectives to achieve consistency and coherence. The experts also had to satisfy themselves that all students had ample opportunity to learn the skills and knowledge necessary to succeed. The statistical results demonstrated that the top scorers in the performance assessments represented a range of student profiles; it was not only the best actors or best readers who achieved top scores.

Criteria-related (or concurrent) validity evidence was studied by comparing the statistical

results with the other data gathered throughout the program. The journal prompts, homework assignments and surveys provided detailed evidence of student learning and interest in the project and showed significant correlation with the statistical evidence where there was a complete data set. A complete comparison of all of the data sources has not been possible due to inconsistency in the collection and scoring of the assignments among the schools. However, the qualitative data provided strong confirmation of the impact on students in the specific categories of Content Understanding, Theater Skills and Learning Behaviors, and gave evidence of that many students were able to integrate their learning through the theatre experience into other areas of school and

their lives. A few samples follow.

My knowledge of Shakespeare as a writer increased significantly. I found that, by the end of the program, I was more able to look at a line of Shakespeare and note not just what the line means but also what techniques Shakespeare uses to convey these meanings. I can more easily spot the puns in Shakespeare, something I used to consider one of the hardest things to do when dealing with Shakespeare.

10th grade student, George C. Marshall HS

Pericles is better understood acting it out, like the rest of Shakespeare's plays, than sitting still reading it. I gained valuable knowledge of how to convey the feelings of the text to the audience and how to interact with the other actors.

12th grade student, Frederick Douglass High School

I know now that Shakespeare is really interesting and is not just something rich white people are into.

10th grade student, Charles H. Flowers HS

In the current research design it is not possible to make claims concerning the predictive validity of

the assessment. An insufficient number of students had taken other tests that would reveal their

knowledge of Shakespeare, such as state regents or advanced placement exams, to allow us to

make statistical comparisons.

Lessons Learned, Next Steps and Future Research

The assessment process has had a major affect on curriculum and instruction in the program. When teaching artists and classroom teachers get clear and immediate information about individual student learning -- as opposed to general impressions about a class as a whole -- their perspectives and approaches change. It is no longer enough to "cover" a topic and move on. Students have to be able to demonstrate their understandings and if it takes more time for all or a large majority of students to develop a skill or gain an understanding then the teacher must reevaluate priorities or reconsider how best and for how long to teach that element. As one teaching artist put it:

The development and implementation of the assessment has allowed us as teaching artists to slow down, have more time for reflection, and make sure that we are truly accomplishing the objectives we've set out to achieve. Before, we were doing great work, but were trying to cram way too much into a single workshop. In developing the assessment, we had to decide what was important – what did we want students to get out of Text Alive! and how would we know if they got it.

The development and observation of the assessment criteria was a major help in clarifying the focus of the program. By defining and writing observable behavioral descriptors for different levels of performance in the six criteria (imagination, content understanding, focus, collaboration, vocal skills, physical skills) the education staff and teaching artists had to concretize learning objectives and come to grips with the fact that students come into the program with different skills and backgrounds and must be assessed fairly and fully regardless of prior experience and their role in the final production. Even students with no prior acting experience, limited interest in theatre, and poor vocal skills must be able to succeed in the program. This awareness helped to shape the balance in the criteria between text understanding, affective behaviors, and theatre skills. We also had to increase the emphasis on stagecraft and other aspects of the production to allow a wider range of outlets for students who were not strong actors. The biggest challenge for teaching artists was to organize the assessment classes to allow every student to be seen and heard and to maintain fairness in the assessment. The teaching artist's role in an assessment context is quite different than in normal teaching; they must carefully monitor the time and feedback given to each student, the distribution and length of student comments and questions, and the natural advantage given to students who perform later in the class and have the benefit of seeing and hearing everyone who came before. The facilitators' instructions and comments have not been scripted, as that would make the assessment process less authentic and engaging for students. It is just as important in theater to see how the actor responds to a suggestion and makes adjustments in the second performance, as it is to see how he or she initially approaches a role. But facilitators must be careful to give the same amount and kind of feedback to everyone, to hold back from the natural tendency to come to the aid of someone who is struggling, and to be constantly aware of time. Discussion and clarification of these issues has been part of ongoing evaluation of the assessment process with teaching artists.

Another issue concerned the necessary reliance on the in-school English and theater teachers to assign and collect the design project and writing assignments that were part of the assessment. Differences among schools in the time allotted to the work, policies and habits toward homework, and follow-though on the part of teachers, posed difficulties in comparing and evaluating the results across schools. At this point, these assignments are considered important curricular elements but require further refinement to be fully incorporated into our assessment process and research. As we clarify the expectations of teachers and continue to refine methods for collecting student work from all of the schools we hope to be able to make better use of a combination of quantitative and qualitative data to investigate the overall impact of the program on various sub-groups including comparisons of English and Theater Arts students, different academic achievement levels, ages and other demographic comparisons.

The main goals of this part of the project -- to study how Text Alive! was working and to use data to improve the program design and instruction – have been substantially reached. We are now prepared to tackle the second goal, which is to conduct empirical research to compare this approach to other ways of helping students understand and appreciate Shakespeare. Such a research study will require commitment and long-term cooperation from the schools. To meet current standards for experimental research in education students would need to be randomly assigned to English classes where they would study Shakespeare with and without the addition of Text Alive! Instruction in the non-Text Alive! classrooms would have to be carefully planned to insure that a valid comparison could be made, both in terms of content and in the opportunity for engagement. We are currently looking for schools to participate in such a study.

Conclusion

Over the past four years, the assessment process has become a basic part of Text Alive! Beyond the initial goal of creating a useful and practical individual student assessment, the process has helped to shape the program design and teaching, providing ongoing, immediate data to help improve curriculum, train and evaluate teaching artists, and evaluate progress in reaching the program's goals. It has also strongly influenced the pedagogy of the teaching artists and classroom teachers involved.

Careful, valid, performance-based individual student assessment is challenging in any situation. It is especially difficult for a program that sees students for a limited amount of time and in highly diverse settings and conditions. It has taken an immense and sustained effort for the education department of the Shakespeare Theatre Company to develop and implement this process. The biggest investment has been in meeting and training time with the staff and teaching artists,

professional development with the partnering English and Drama teachers, and the ongoing commitment of program assistants and interns to support the assessment and data management. But the payoff in terms of direct information about the learning progress of each and every student involved in Text Alive! provides powerful evidence for the benefits of assessment to improve teaching and learning in a theatre residency program.

		ldeal	Excellent	Good	Adequate	Needs Support
		<u>5</u>	<u>4</u>	<u>3</u>	<u>2</u>	<u>1</u>
Behavior Skills (Focus and Cooperat ion)		1. Participates actively, helps class maintain focus and shows leadership by taking risks, trying new things, volunteering, and expressing positive energy.	1. Participates actively and maintains focus in all activities.	1. Participates actively and usually maintains focus in most activities.	1. Participates actively in some activities and struggles to maintain focus.	participate
	g, respond ing,	2. Actively listens, observes and follows all directions. Asks good questions to clarify instructions; helps other students listen, observe and follow directions.	listens to and observes directions and other students. Asks	other	2. Inconsistently listens to and observes directions and other students.	2. Does not listen or pay attention to directions or other students.
Imaginati on / Content Understa nding	ation,	1. Generates and offers appropriate, unique, personal ideas and insights freely and fully.	1. Generates and offers ideas freely and fully.	1. Generates and offers some ideas.	1. Generates and offers few ideas.	1. Does not generate or offer ideas.
	Underst ands	2. Expresses deep understanding and insights and probes more deeply into text through personal choices and verbal responses	2. Clearly and insightfully shows interpretation of text through verbal response or strong personal choices.	verbal	interpretation	2. Shows no interpretation of text through verbal response or personal choices.

Theatre Skills	Vocal	1. Speaks with complete command, authority, and clear intention which illuminates and communicates the words being spoken	1. Speaks loudly and enunciates words clearly and confidently.	1. Speaks loudly and enunciates clearly enough to be understood.		1. Speaks softly and is difficult to hear and understand.
	Physica I	2. Shows outstanding command of the body through detailed and believable embodiment of characters or states of being while showing sensitivity and awareness to others and the surroundings. Uses stage space to effectively portray relationships, states of being, or ideas.	2. Responds with whole body while demonstrating acute and accurate awareness of physical space and other people.	2. Responds with whole body while demonstrating awareness of space and other people.	2. Demonstrates limited physical responses and limited awareness of space and other people.	2. Demonstrates extremely limited physical responses and no awareness of space and other people.

Scoring Rubric for Shakespeare Theatre Company Observational Assessmen