

Evaluating a Teacher Education Program Through Performance Task Assessments

Sara Ashworth, Florida Atlantic University

Megan Dolan, Florida Atlantic University

Gary Kirsch, Florida Atlantic University

Michael Vitale, East Carolina University

Introduction

A highly qualified teacher in every classroom. It has become the latest battle cry for school reform at the national, state and local levels as the growing research base on teacher effectiveness supports the notion that student learning and achievement is closely linked to teacher quality (Darling-Hammond, 1997; Haycock, 1998; Sanders & Rivers, 1996). While more and more communities struggle with current and impending teacher shortages and the public continues to lobby for smaller class size, educational and political leaders are faced with the difficult challenge of ensuring that every student has a qualified teacher, a task now mandated by the federal *No Child Left Behind* legislation. Within a developing landscape of national accountability in which the quality of teacher education programs is constantly under fire, it is critical to conduct ongoing systematic evaluation of the degree to which students absorb knowledge in such programs with the specific goal of improving the preparation of future teachers.

The research on teacher preparation focuses on a number of issues. Some studies are concerned with which types of programs (traditional college of education programs or alternative certification programs) best prepare pre-service teachers. Additional studies have focused on the types of classes pre-service teachers should be taking, and whether a greater emphasis should be placed on pedagogical courses or content courses (Ballou & Podgursky, 2000; Darling-Hammond, 2000; Druva & Anderson, 1983; Rigden, 2002). Other studies look at whether teacher quality improves when the teachers have advanced degrees (Goldhaber & Brewer, 1997; Rowan, Correnti & Miller, 2002). In investigating these topics, researchers have generally relied on data concerned with prospective teacher perceptions, pre-service teacher portfolios, and compliance with licensure and NCATE requirements (Darling-Hammond, 2000; Gitomer, Latham & Ziomek, 1999; Goodlad, 1990). These types of studies encounter some significant limitations, however. State licensing exams, for example, rarely include performance tasks but instead rely on multiple-choice questions, which call for the student to recognize, rather than generate, a correct answer. Perception data can be valuable in gauging what students thought or felt about a specific teacher education program, but are not necessarily reliable measures of the actual quality of that program. As the stakes on accountability for teacher preparation programs continue to grow, it is clear that a different kind of measure will be needed (French, 2002).

More recently, consistent with the national trend toward increased accountability, researchers are striving to correlate different types of teacher preparation directly to the achievement of those new teachers' students. These studies, however, must control for the plethora of other variables that also impact student learning and performance on standardized tests (Wilson, Floden & Ferrini-Mundy, 2002). The research conducted in this area is scattered, and often inconclusive (Allen, 2003). While correlating teacher preparation with student achievement is an admirable goal, for it is indeed student achievement that is our ultimate concern, there must be an intermediate step that provides a better picture of the quality of teacher preparation, while relying on more accurate measures than multiple choice tests, portfolios and perception surveys.

This study investigated the effectiveness of a teacher preparation program by asking participants to complete specific performance tasks representing basic pedagogical knowledge that was repeatedly reinforced throughout their core teacher education courses. In contrast to traditional multiple-choice assessments, which require students to *recognize* correct answers, these tasks required the students to *recall* what they had learned. The tasks measured the degree to which the program students had absorbed and retained the knowledge presented in their coursework, by demonstrating their abilities to produce an educationally sound lesson plan, including an assessment plan, higher and lower order questions, and specific academic praise statements. Participants were also asked to rate their perceptions of the teacher education program, and these ratings were compared to their performance on the tasks.

Sample

The focus of this study was a teacher education program at a large university in the Southeast, with 23,000+ students over seven campuses. 199 prospective teachers at three campuses took part in the study at the start of their student teaching experiences during the spring of 2003. This followed two pilot studies in 2002 and 2001, with samples of 250 and 280 respectively, during which the survey instrument and implementation procedures were refined. The current sample group included 166 students in the Elementary Education program, 11 students in the Secondary Education Program and 17 students in the Exceptional Student Education Program (five students did not identify their program).

Procedures and Instrumentation

Surveys were administered to three groups of students on separate campuses during a special hour-long session prior to their student teacher meeting. Directions were read from a predetermined script, to ensure that each group received the same instructions. The survey included both performance tasks and perception questions. For the performance tasks, participants were asked to write out a lesson plan similar to those they were required to do in each of their core course. They were asked to identify a specific concept in Language Arts/Reading, Math or Science (randomly assigned for elementary education majors, chosen by the participant for secondary education majors), and state the overall instructional objective for the lesson. They were then asked to provide an initiating activity, three core activities, and an assessment plan (including the specific content to be assessed and the plan of action). In addition, they were asked to provide four lower order questions, four higher order questions and four examples of specific academic praise appropriate to use when teaching this concept. The perception questions used 5-point Likert scales and called for participants to gauge how prepared they were overall and by their individual courses during their College of Education experiences. The survey also asked them to state the degree to which they felt confident to enter the classroom.

Analysis

Participant responses were evaluated in two stages based on comprehensive criteria in keeping with the state's in-service teacher evaluation criteria, the standards for accomplished practices, the state K-12 academic standards and the state K-12 standardized tests specifications. First, responses were rated as appropriate or not. For example, was the identified concept appropriate for a lesson, did the initiating activity clearly introduce the new concept, did the higher order question call for the student to analyze, synthesize or evaluate, etc. Second, the

responses were categorized by type. For example, did the lower order question represent knowledge, comprehension or application? The survey responses were tallied by one researcher with more than 34 years of active educational experience using a detailed criteria matrix refined from the previous pilot studies. To verify reliability, a random selection of surveys was tallied again by a second researcher with extensive practitioner experience.

Findings

Performance Task Questions

In the first stage of data analysis, responses were evaluated based simply on whether or not they met the minimum criteria for an appropriate answer. Figure 1 shows the percentage of appropriate responses for each performance task. As the data were analyzed, several qualitative categories emerged among the responses, and each response was subsequently evaluated as to which category it represented. This evaluation revealed some interesting patterns that provide a more robust picture of the participants' responses.

Of the 199 students surveyed, 29% stated a specific academic concept, based on the state's criteria for in-service teacher evaluation, for example, "As an ongoing part of the human body section: skeleton we will be learning about joints: immovable, hinge, ball & socket." 62% provided responses that were actually tasks rather than concepts. Responses included "telling time in five-minute intervals," "multiplication of single digit number," and "Nouns." When asked to state the instructional objective for the lesson, 24% gave responses that reinforced the overall learning experience,

while 44% of the responses were essentially the same as the concept stated. On one survey the concept stated was "I intend for the students to identify the greater number," and the objective was "the student will be able to identify the greater number."

The survey called for participants to describe an initiating activity and three core activities they used (or could use) to teach the stated concept. 30% of the participants identified an introductory activity that would prepare their class to learn a new concept, while 58% described activities that were not distinguishable from standard core activities. An additional 10% simply wrote the name of a general task, such as KWL chart, worksheet, read a book, etc. Of the core activities provided, only 29% were directly relevant to the lesson concept. 66% of the participants provided a set of three activities that were in effect the same activity reiterated three times. One participant described three activities for single digit multiplication, "1) The student will answer problems in math book on calculating single digit math problems. 2) The student will pair off and answer as many math problems within five minutes using flash cards. 3) The student will watch the flash card video and answer the questions at the rate of the speaker."

The surveys were divided by subject area. There were 67 in Language Arts/Reading, 66 lessons in Math, and 66 in Science. While there were no patterns noted among subject areas in

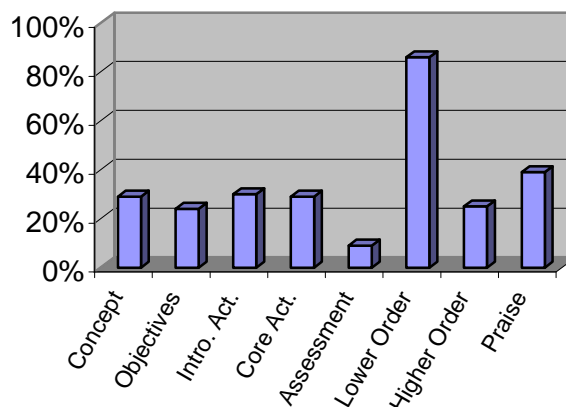


Figure 1. Percentage of Appropriate Responses

assessment, questioning or praise, there were some noteworthy findings in the lesson plans. Of the lesson plans rated as having activities that reinforce the overall content, there were 12 in Language Arts/Reading, 16 in Math, and 27 in Science. As for initiating activities that genuinely introduced new content to the students, there were 14 in Language Arts/Reading, 17 in Math, and 30 in Science.

For the assessment portion, the survey gave the following instructions, “Describe the assessment plan you used (or could use) to determine that this concept was learned. Indicate the specific content to be assessed and plan of action.” Only 9% gave responses that specifically described how the assessment method would assess the students’ overall learning of the concept presented, while 69% of the participants simply offered a type of assessment (KWL, test, practice sheet, observation, completion of homework etc.). 17% wrote general statements such as “See how much they know in math.”

86% of the responses for lower order questions called for one of the first three cognitive levels of Bloom’s Taxonomy, as they should have. However, it is interesting to note that 91% of these responses represented knowledge questions (the lowest level of the taxonomy), while 8% represented comprehension questions, and only 2% represented application questions. 93% of the questions would require a short 1-3 word answer. When asked for examples of higher order questions, barely 25% of the responses represented analysis, synthesis or evaluation, and 55% of the responses were actually lower order questions. 25% called for the student to “compare or contrast,” while 28% asked for the students’ opinion.

Regarding specific academic praise, 39% of the responses included both a positive value statement and a specific academic reference (as required by the state’s in-service teacher evaluation criteria). 45% of the participants provided examples which were from the viewpoint of pleasing the teacher, using phrases such as “I like the way you...” or “I’m impressed...” to initiate the praise statement. 19% of the responses dealt with behavioral elements rather than academic elements, for example “I like the way you are following directions,” or “Your group worked well together.” 55% of the participants included a student’s name in their example.

Finally, each survey was given an overall score representing the percentage of performance task responses that were evaluated as appropriate out of the total number of responses requested. The average score was 41%. There were no participants who completed all of the questions satisfactorily, and only one fourth scored higher than 50%.

Perception Questions

Participants we asked to rate the degree to which they were prepared by the teacher education program for their student teaching experience. 24% indicated they were “very well prepared,” 54% indicated “mostly prepared,” 17% selected “somewhat prepared,” and 2% chose “minimally prepared.” None of the participants selected “not at all prepared.” Figure 2 details these findings.

A final question was included, in which participants were asked to gauge their confidence

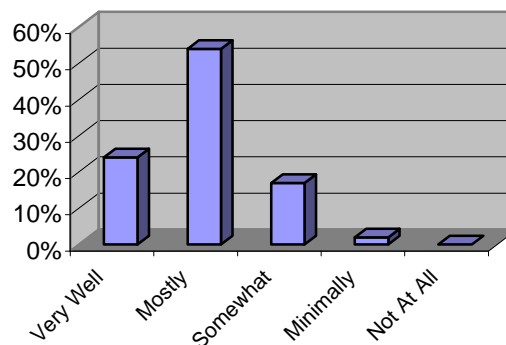


Figure 2. Degree to which participants felt prepared to teach

in their present ability to perform as a teacher in a classroom setting. 25% responded that they were “highly confident,” 57% indicated “confident,” 16% selected “somewhat confident,” 1% indicated they were “not confident,” and 2% chose “not prepared.” Figure 3 shows these results

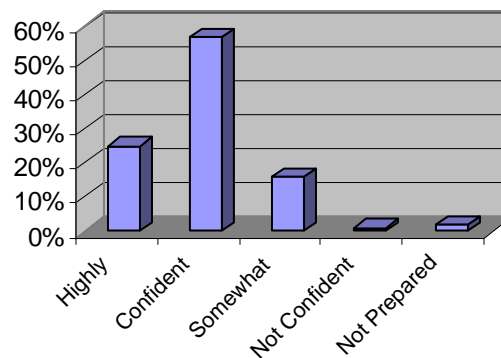


Figure 3. Degree to which participants felt confident to enter the classroom.

Additional Analyses

Multiple regression analyses were conducted to determine if there were correlations between student performance on the survey questions and their primary campus location, their perceptions regarding their preparation, and their degree of confidence in entering the classroom following their student teacher experience. There were no significant correlations found for any of these categories.

Discussion

The purpose of this study was to assess student learning in a teacher education program and to investigate alternative methods for assessing this learning to the ones currently employed by the department (perception surveys, portfolios, and recall-dependent multiple choice tests). Additionally, the researchers set out to determine what relationship existed between the perceptions and abilities of the pre-service teachers in this program. Within the developing context of educational research, in which a greater emphasis is being placed on scientifically based studies (Shavelson & Towne, 2002), if one is seeking to validate the retention of knowledge within teacher preparation programs, it follows that one should rely on measures which go beyond participant perceptions.

The teacher education department in question has a professional mission to provide “quality programs for prospective teachers, in-service teachers, curriculum coordinators and other school personnel.” One would therefore hope that the program would at least provide the basic knowledge, the habit of mind for instruction, that a classroom teacher needs and more specifically the knowledge that is identified in the syllabi. From an in-depth review of the core course syllabi, one would assume that the students in this program are exposed to an excellent curriculum of pre-service teacher experiences, but the data gathered in this study paints a disturbing picture of the knowledge actually absorbed by the students.

Most of the lessons plans were designed around a task or skill, indicating that the students did not think conceptually about their instruction. Current educational research supports the idea that students learn best when their lessons are conceptually organized around big ideas, rather than random assortments of discrete activities (Bransford, Brown & Cocking, 2000). In fact, this element of the survey caused a great deal of debate among the researchers in determining which responses were indeed concepts and which were not. This particular analysis relied on Merrill and Tennyson’s definition of a concept as “a set of specific objects, symbols, or events which are grouped together on the basis of shared characteristics and which can be referenced by a particular name or symbol” (1977, p.3). Ultimately, the researchers found that for the purposes of this study they could not reach a conclusion simply by reading the statement of concept. Instead they found that the statement of concept had to be considered along with the activities

included in the lesson plan. By reading through the entire lesson plan, the researchers could then verify how the participant understood the specific concept. This may indicate that individual faculty members vary in their ideas of what defines a concept, or that the learning of the definition occurs in isolation rather than in application while developing a lesson plan. The findings regarding the instructional objectives were no less disturbing. Almost half the objectives provided were in effect identical to the stated concept for the lesson. This suggests that these participants are unable to distinguish between a lesson concept and objective. These are key issues if prospective teachers are going to learn to develop quality lessons that reflect our current understanding of the best ways in which students learn new knowledge.

The analysis of the lesson plan activities prompted a number of observations. More than half of the examples provided as initiating activities did not serve to initiate the students to a new concept, and were in fact no different from other curricular activities. It would appear that these participants are unclear about the purpose and characteristics of an initiating activity. An overall evaluation of the lesson plan found that only one fourth of the plans included complementary activities which supported learning of the instructional objectives. One third had sets of activities listed that were essentially a reiteration of each other, but did at least support the intended concept, while another third had sets of tasks that did not support the intended concept. The researchers also examined the lessons to gauge how high the learner engagement was for each activity. About one third of the lessons prompted high learner engagement, about one third prompted low engagement, and about one third were mixed. Wiggins and McTighe (1998) describe an “activity driven curriculum,” in which the activities are chosen and arranged based on factors other than student learning, and many of the surveys represented this type of lesson. The participants seemed to be looking for activities that were clever, creative, or fun, which on the surface is admirable, but failed to fully examine the learning outcomes the lessons evoked. Only one fourth of the lessons featured a thoughtful progression of connected activities that had educational value and would support the learning objectives.

One of the most difficult tasks for a teacher is designing appropriate assessment tasks. An important aspect of this is ensuring that the task is a valid measure of the knowledge or skill to be assessed. Although the survey called for a specific “plan of action,” 65% of the participants only listed an assessment technique, without elaborating as to how it would assess student learning of the specific concept. It was impossible to determine if the technique was appropriate or not. Wiggins and McTighe, in their “backward design” model, recommend that teachers should determine what the assessment is before choosing the activities (1998). This helps to ensure that the activities result in the instructional objectives to be measured.

The lower and higher order questions revealed that most of participants were very capable of sculpting questions that called for a factual, knowledge answer. Very few, however, wrote questions that went beyond knowledge into comprehension or application. It is also noteworthy that more than half of the questions the participants indicated as higher order were in fact lower order, suggesting that these pre-service teachers do not have a good understanding of the distinction. As current research emphasizes the importance of developing deep understanding by promoting higher order processes (Bransford, Brown & Cocking, 2000), it is imperative that future teachers are able to do so. It is also a critical issue for teachers in this era of accountability, as 30-70% of the questions on the state standardized tests for elementary and secondary students are higher order in nature (the test includes progressively more higher order questions as students advance in grade level).

Another area that has been the subject of educational research is the effect of different kinds of praise on student achievement. This research has yielded some valuable insight into the attributes of effective praise, and recommends that praise specify the nature of the accomplishment and reward attainment of particular performance criteria (Brophy, 1981). The state's in-service teacher evaluation process includes a component focusing on teachers' use of specific academic praise. It is unsettling that only 72% of the participants were able to provide any examples of specific praise, and only 7% were able to provide four examples, as the survey asked. More than half of the participants included a student's name in their example, as if this made the reference "specific."

One of the most profound discoveries of this study was the disparity between participant perception and performance. While 78% of the participants indicated they felt mostly or very well prepared by their College of Education courses, and 82% indicated they felt confident or highly confident in their present ability to perform as a teacher, there were only two categories of performances tasks (lower order questions and specific academic praise) in which more than 30% of the responses were appropriate. In all of the other categories regarding lesson concepts, objectives, introductory activities, core activities, assessment and higher order questioning, participant performance was very poor, with 30% or less of the responses meeting the criteria. On average, participants provided acceptable responses in less than half of the categories. When the overall performance of participants is contrasted with their perceptions regarding preparation and confidence, a stark contrast appears.

Conclusions

The results of this evaluation were rather bleak. While the richest data was found qualitatively through the actual examples of the performance tasks, there was much to be learned from the descriptive statistical findings. Although the students were asked to construct lesson plans in all of their core teaching courses, they were largely unable to develop coherent, logically sequenced plans geared toward specific academic learning goals. While most of their courses addressed cognitive processes and critical thinking, only a small number were able to adequately produce higher order questions. Very few were able to describe appropriate assessment activities related to their instructional objectives, yet most thought that they had done so. While the current findings are limited to students from one semester, the data from this sample are remarkably consistent with the data from the two previous pilot studies.

From these data the researchers came to several conclusions. First, students in this department do not appear to be receiving adequate instruction on how to develop a lesson plan, or are not receiving adequate feedback on the ones they do develop. This includes the quality of the lesson concept, the clarity and appropriateness of lesson objectives, the capacity of the initiating activities to introduce the students to a new concept and activate prior knowledge, and the capacity of the core activities to support the instructional objectives and overall learning experience (including how they are sequenced). This cannot be stressed enough. If K-12 students are going to achieve an optimal level of knowledge and skill, their teachers must, *at a minimum*, be able to design quality lesson plans that support conceptually organized learning. Instruction and feedback need to emphasize both content and pedagogy. It should be noted that in the section of the survey reserved for participant comments, 68 students suggested the need for more instruction in classroom management. It is possible that some of the issues they were beginning to encounter in their student teaching experiences, which they noted as classroom

management concerns, were actually connected to their inability to prepare a thoughtful, well-sequenced lesson plan.

Second, program students need more specific instruction and practice in developing valid assessment plans for their lessons. While there is clearly a growing emphasis on large scale standardized assessments in education, there is also a growing recognition of the importance of classroom level formative and summative assessments (Pellegrino, Chudowsky & Glaser, 2001). Instruction on the development of assessments must be closely tied to instruction on the development of lesson plans, so that prospective teachers can see the crucial interrelation of these important elements.

Third, some consensus must be established among the faculty (and the educational community in general) regarding key educational ideas and terms. The issue of diverse definitions and understandings is one that plagues all of education today, but is particularly critical in developing new teachers. Judging from the surveys, there is wide disparity among the faculty regarding definitions of the following terms: *concept*, *instructional objective*, *initiating activity*, *assessment plan*, *lower order*, *higher order* and *specific academic praise*. Consensus about these terms is important so prospective teachers will have a clear understanding of these key ideas, and because they figure prominently into their future in-service teacher evaluations.

Fourth, a *consistent and high standard* of work and performance must be established and uniformly implemented throughout the teacher education program. While the intended curriculum (as reflected by the syllabi) may be adequate, the actual curriculum (what is happening in the classroom) is not achieving the intended results. What is perhaps most distressing about this study is the fact that most of the participants felt at least well prepared and confident to enter the classroom, despite their inability to develop a quality lesson plan, ask higher order questions, or construct effective praise statements. Most of the participants believe they know what they are doing. Research on misconceptions suggests that it is much more difficult to change a person's misunderstanding of conceptual knowledge than it is to promote correct understanding in the first place (Schneps & Sadler, 1998). As these teachers become more set in their flawed routines regarding lessons, questioning and praise it will be less likely that they will be able to learn effective routines in the future. It is vital that future teacher education students receive accurate critical feedback, both to correct misconceptions and to support more valid assessments of their abilities.

Finally, the researchers recommend that this department continue to explore the use of performance tasks (both in their courses and as summative assessments) to determine how well their students are absorbing the critical knowledge intended by the core course curricula. This study has cast considerable doubt on the validity of perception data as a measure of student learning. It has also raised questions as to the validity of portfolio data (or perhaps the evaluation of that data). If the students are using lesson plans similar to those created for this survey as part of their portfolios, and are still judged as ready to teach, there may be an issue regarding how these portfolios are assessed. By relying instead on performance task data, such as was gathered in this study, department stakeholders can get a more accurate picture of what their students are learning, and more importantly, how prepared they really are to teach future generations of children.

This study focused on a specific teacher education department, so the findings that detail the effectiveness of this program cannot be generalized to other programs. However, the wide disparity between perceived abilities and demonstrated abilities on the basic tasks selected suggests that there may be significant implications for other teacher education programs

regarding how they evaluate their performance. Based on perception data, portfolios, grade point averages and rates of graduation, certification and employment, the faculty of this program believed they were providing an excellent preparation experience for their students. If other programs are relying on this same kind of data, they may be drawing the same erroneous conclusion. Aside from the recent emphasis on accountability, the impact of teacher quality on student learning demands that teacher education programs accurately measure the degree to which their students are being prepared. This study provides a model on which other programs can base their own evaluations.

References

- Education Commission of the States (2003) *Eight questions on teacher preparation: What does the research say?* Denver, CO: Allen, M.
- Ballou, D. & Podgursky M. (2000) Reforming teacher preparation and licensing: What is the evidence? *Journal of Teacher Education, 102*, 5-27.
- Bransford, J., Brown, A. L. & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.
- Brophy, J. (1981) Teacher praise: A functional analysis. *Review of Educational Research, Spring, 26*.
- Cochran-Smith, M. (2002). The research based for teacher education: Metaphors we live (and die?) by. *Journal of Teacher Education, 53*, 283-285.
- Darling Hammond, L. (1997). The quality of teaching matters most. *Journal of staff development, 18*, 38-41.
- Darling-Hammond, L. (2000). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record, 102*, 28-56.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*.
- Druva, C.A., and Anderson, R.D. (1983). Science teachers' characteristics by teacher behavior by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching, 20*(5), 467-479.
- French, R. (2002). Linking teacher quality to student achievement through assessment. *Basic Education: A Monthly Forum for Analysis and Content, 46, 10*, 12-16.
- Gitomer, D.H., Latham, A.S., and Ziomek, R (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Educational Testing Service.
- Goldhaber, D.D., and Brewer, D.J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources, 32*(3), 505-523.
- Goodlad, J.I. (1990). The institutional and regulatory context. *Teachers for our nation's schools* (pp. 108-153). San Francisco, CA: Jossey-Bass.
- Haycock, K. (1998). Good teaching matters: How well-qualified teacher can close the gap. *Thinking K-16, 3*, 1-14.
- Merrill, M.D., & Tennyson, R.D. (1977) *Teaching concepts: An instructional design guide*. Englewood cliffs, NJ: Educational Technology Publications.

- Morgan Quitno (2003) *Education State Rankings*. Retrieved October 15, 2003 from <http://www.morganquitno.com/edrank.htm>.
- Pellegrino, J.W., Chudowsky, N. & Glaser, R. (2001) *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Rigden, D. (2002). The STEPS that support P-12 learning and achievement. *Basic Education: A Monthly Forum for Analysis and Content*, 46, 10, 1-4.
- Rowan, B., Correnti, R., and Miller, R.J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record*.
- Sanders, W.L. & Rivers, J.C. (1996) *Cumulative and Residual effect of teacher on future student academic achievement. Research Progress Report*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Schneps, M.H. & Sadler, P.M.(1987). *A Private Universe*. Harvard-Smithsonian Center for Astrophysics, Science Education Department, Science Media Group, Washington, DC: Annenberg.
- Shavelson, R.J. & Towne, L. (2002). *Scientific Research in Education*. Washington, DC: National Research Council.
- Wilson, S.M., Floden, R.E. & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*, 53,190-204.