Included in the Pennsylvania State University teacher education program for secondary school teachers is a requirement that students demonstrate ability to read, write, and speak effectively in professional situations. The basic assessment procedure in these skills is conducted before the student teaching experience. The assessment includes three parts: (1) the Nelson-Denny Reading Test; (2) the Criterion-Referenced Test of Reading/Writing Competence (CRT); and (3) a videotaped speaking assessment. The development and implementation of these procedures is described, and information is given on: (1) how the Nelson-Denny Reading Test is administered and scored; (2) reading objectives established by the CRT (understanding the professional vocabulary, ability to answer literal level comprehension questions, ability to answer inferential level comprehension questions, and ability to interpret information from tables); (3) the rating system used in scoring the CRT; (4) the methods of rating and evaluating the videotaped speeches; and (5) the results of the assessments and evaluation of their reliability and validity. Some discussion is included on the test scores of the university's prospective teachers. A discussion is also presented on conclusions drawn from the demonstrated reliability and validity of the assessment procedures. (JD)
The Basic Skills of Prospective Teachers: How Well Do They Read/Write/Speak?

Mary M. Dupuis
Edward R. Fagan
The Pennsylvania State University

Presented at:
American Educational Research Association Conference
Montréal, Quebec
April 15, 1983
The Basic Skills of Prospective Teachers: How Well Do They Read/Write/Speak?

Mary M. Dupuis and Edward R. Fagan
The Pennsylvania State University

The popular press has been reporting in the past few years that teachers are not well skilled in areas which they are called upon to teach (Leiser, 1981; Lyons, 1980). Educators have long accepted that a teacher cannot teach what s/he does not know. This syllogism seems especially telling when it is applied to the basic skills of reading, writing, and speaking. With recent public attention focused on basic skills instruction, the question of teachers' level of competence in these basic skills becomes a matter of some importance to teacher educators.

The Penn State University faculty preparing secondary teachers became concerned about this problem several years ago. At issue was the question: how can we verify to ourselves and to potential employers of our graduates that they are competent in these three basic skills? A search of the literature was instituted looking for effective assessment procedures and evidence of teachers' ability (or inability) to use these basic skills. In fact, we found no evidence of teachers' basic skill levels, either positive or negative, except for a small study of elementary teachers' ability to use study skills (Askov, Kamm, and Klumb, 1977). We questioned, then, how effectively our prospective teachers could use their communication skills in professional situations.

The authors are grateful to Sandra L. Snyder, Mary Kopa, and Bernard J. Badiali for their assistance in conducting this research.
To clarify these basic skills, a Penn State faculty has agreed that reading, writing, and speaking "basic." Although we can add a number of skills we believe to be "basic" to good teaching, we have agreed this generic definition of the basics. We have more recently added computation to these basics, but we do not report on that skill here.

A little background. The Penn State program in Secondary Education prepares teachers in English/Communication, foreign languages, social studies, mathematics, and the sciences. A newly revised program, under development for over two years, was formally begun in September, 1979. During the period of revision, the program was formulated on a competency base. Included in this list of competencies was a straightforward statement that each student will demonstrate his/her ability to read, write, and speak effectively in professional situations.

When the program faculty became serious about making this competency operational, it was easy to decide that the University's introductory coursework in these skill areas was not sufficient to verify professional level competence. That is, requiring a grade of C or better in freshman composition or speech courses was not sufficient. The development and implementation of the necessary assessment procedures is the subject of the paper which follows. Our goal is to demonstrate to a skeptical public--and ourselves--that the teachers we recommend for certification can use these three basic skills effectively in their professional work.

The Assessment Procedure

The basic assessment procedure is conducted in the first of three generic methods courses preceding student teaching. A full description
of the Penn State program is available in Fagan (1981). The assessment procedure includes three parts: (1) The Nelson-Denny Reading Test, (2) The Criterion-Referenced Test of Reading/Writing Competence, and (3) a videotaped spelling assessment.

Assessing our students' reading ability was a major concern, because there is no University-wide assessment of reading, hence no data of any kind on students' reading levels. As a result, we included two different assessments of reading. The Nelson-Denny Reading Test Form D, a widely used standardized test for adults, was chosen because it has demonstrated validity and reliability (.95 at the college level). Furthermore, it is easy to administer and score. We use the short version, which allows 10 minutes for the vocabulary section, and 20 minutes for the comprehension section. The 30-minute test was administered during a regular class meeting. Three scores are generated from the Nelson-Denny: Vocabulary, Comprehension, and Total. The Total score is figured from the raw scores according to the formula $T = V + 2C$. Percentile scores are derived for each of the three raw scores.

The Criterion-Referenced Test for the Assessment of Reading and Writing Skills of Professional Educators (called the CRT) was developed as a test of professional reading and writing skill. Whereas the Nelson-Denny uses passages on general topics for its comprehension assessment, the criterion-referenced test is based on the reading of an article in a professional journal, "What's New in Ability Grouping?" (B. J. Wilson and D. W. Schmits) from the Phi Delta Kappan (59:8, April 1978, 535-6). The rationale for such a professional reading source is that a teacher should be assessed on reading material similar to that which s/he will be
reading as part of his/her continuing professional development. The 
Kappan is recognized as a respected and well-written journal of general 
interest to teachers.

The criterion-referenced test establishes four objectives to be measured for reading:

Given the professional reading selection, the prospective teacher will demonstrate:

1. understanding of professional vocabulary used in the reading;
2. ability to answer literal level comprehension questions (known as low comprehension);
3. ability to answer inferential level comprehension questions (known as high comprehension);
4. ability to interpret information found in tables.

Thus, the test has four sections and we derive five scores from it:

1. Vocabulary -- 6 items
2. Low Comprehension -- 5 items
3. High Comprehension -- 9 items
4. Data Interpretation -- 3 items
5. Total -- 23 items

Comprehension levels here are defined as in Barrett's Taxonomy of Reading Comprehension (Barrett, 1972). The four levels given by Barrett correspond roughly to Bloom's Taxonomy of Cognitive Levels (Bloom, 1956) for Levels 1, 2, and 3 (see Figure 1), while Level 4, Appreciation, belongs in the Affective Domain. The CRT in reading tests only Levels 1 and 2 of Barrett's Taxonomy. Level 3, Evaluation, is tested by the writing sample, a second part of the CRT. For the writing sample, students are asked to respond in writing to one of two Level 3 questions related to the same reading selection:
1. What does the information contained in this article suggest about the influence of research results on instruction? In your opinion, is this an accurate representation of the application of research findings in instructional practices in general? Suggest ways that research could have a greater influence on instruction.

2. The author notes that the data reveals a "consistency agreement among teachers" regarding ability grouping. Suggest some possible reasons for this widespread agreement.

This organization for assessing reading is based on the need for readers to move up the taxonomic levels as they read and consider a passage. Thus, our assessment procedure asks the prospective teacher to (1) read the article; (2) complete the CRT on reading; (3) complete the writing sample. The entire assessment can be completed by most students in a 75-minute period, although the test is not timed. Students are allowed to use as much time as necessary to complete the assessment which is given in our Instruction Support Center, an independent learning center, at which the student makes an appointment at his/her convenience.

These two assessments of reading clearly have several differences:

- criterion-referenced vs. norm-referenced
- not timed vs. timed
- individual scheduling vs. group testing
- professional material vs. general material

The reading assessments are computer scored, yielding raw and percentile scores for the Nelson-Denny, and raw scores for the CRT.

The writing assessment is scored holistically, as prescribed by the College Entrance Examination Board (Kirrie, 1979). In this study, holistic scoring followed the procedures given in Cooper (1975), with scores given from 1 (low) to 4 (high).
Figure 1
Barrett's Levels of Comprehension Related to the Taxonomy of Cognitive Levels

<table>
<thead>
<tr>
<th>Bloom</th>
<th>Barrett</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Evaluation</td>
<td>1 Evaluation</td>
</tr>
<tr>
<td>5 Synthesis</td>
<td></td>
</tr>
<tr>
<td>4 Analysis</td>
<td>2 Inferential</td>
</tr>
<tr>
<td>3 Application</td>
<td>3 Literal</td>
</tr>
<tr>
<td>2 Comprehension</td>
<td></td>
</tr>
<tr>
<td>1 Knowledge</td>
<td></td>
</tr>
</tbody>
</table>
Because of the volume of essays, several raters are required. The total N of 217, covering five separate cycles of students, means that 434 essays were read (each essay was read by two separate raters). In addition, the Secondary Education program is an ongoing process running two separate assessments of new groups of prospective teachers each year. Thus, corollary concerns were the establishment of adequate interrater reliability and the training and use of different raters over time.

The first essay rating session in Fall 1980 (using 43 essays) used nine raters (one faculty member and eight graduate assistants). After a two-hour training session, the raters' reliability was figured at .71 using Winer's (1962) formula. The second rating session in Fall 1981 (using five essays) used six raters (two faculty members and four graduate assistants). After similar training, the interrater reliability was figured at .73. Only two of the fifteen raters were the same for both sessions. The 1982 raters (three faculty and five graduate assistants) reached an interrater reliability of .80. We believe that both conditions stated were met: the interrater reliability was high enough for our assessment purposes, and raters from different backgrounds were trained in a reasonable period to conduct the ratings. Among the raters in these sessions were professionals from reading and English, as expected, but also from science, foreign languages, math, and social studies.

Oral speaking has received less attention in the public press and the professional discussions of communicative competence than reading and writing. However, teacher educators generally agree that oral communication is used more frequently in the classroom than writing. Research in teacher effectiveness underscores the importance of oral communication (Snyder, 1981b). Speaking assessment in this program is designed to place
prospective teachers in a professional situation, speaking in a fairly formal way to their peers. Groups of four to six prospective teachers from different content areas speak on a professional topic of their choice for five minutes. These speeches are videotaped so they can be assessed at a later time by competent raters. These videotapes can also be viewed by the speaker for self-evaluation and, if necessary, to provide the basis for diagnosis and remediation.

The scale used in rating these speeches was developed as part of the assessment program. The development and validation of the scale is reported in detail elsewhere (Snyder, 1981a, 1981b). The scale (given as Figure 2) contains five areas, each one containing specific components. Each rater checks the components occurring within a speech, then converts the checkmarks to a numerical score from 1 (low) to 4 (high) on each component. The components are weighted according to the importance of that component to the speaking situation. Weightings are derived from both speech communication and teacher effectiveness research (Snyder, 1981b). The greatest weight is placed on Organization and Development and Adaptation to the Audience. These two components account for nearly half (48) of the 100 points possible. A sixth component, Overall Impression, provides for a holistic evaluation of the speech, using the same procedures as the holistic assessment of writing.

Each speech is rated independently by two raters after a two-hour training session. As with writing, the training session for speaking follows CEEB suggestions and uses speeches drawn at random from earlier assessments. The first set of speeches in Fall 1980 (n=44) was rated by
Figure 2
Speaking Assessment

RATING FORM

Key:
1 - poor
2 - below average
3 - above average
4 - very good
✓ - each descriptor where performance is acceptable

Organization and Development

1  2  3  4  (x6)

Purpose clear
Main ideas clear
Main ideas consistent with purpose
Smooth transitions

Adaptation to Audience

1  2  3  4  (x6)

Provides sufficient information
Eye contact with audience
Relate message to audience
Clear explanations

Language Usage

1  2  3  4  (x5)

Use of appropriate vocabulary
Use of standard English dialect
Enunciation
Use of conventional grammar

Ability to Motivate Audience

1  2  3  4  (x3)

Personal involvement
Speaks expressively
Uses variety in presentation
Uses visual aids

Delivery

1  2  3  4  (x3)

Speaks audibly
Speaking rate
Posture
Body movement and gestures

Overall Impression

1  2  3  4  (x2)

Total score

Comments:

Developed by Sandra Snyder
four raters, all graduate assistants. Their interrater reliability (Winer, 1962) was assessed at .92. The second set of speeches in Spring 1981 (n=55) was assessed by seven raters, all but one of whom were different from the first set. The interrater reliability of the second assessment was figured at .93. The most recent training session (Fall 1982) included eight raters (four faculty and four graduate assistants) rating ten speeches (10 percent of 97). The interrater reliability was .91. Both the reliability and the training of multiple raters meet the program’s criteria for precision, consistency, and practicality.

Results and Their Implications

We have looked at the results of these assessments in two ways, related to the two questions given earlier. First, we considered their reliability and validity. We evaluated the results of the reliability studies reported above and determined that the scoring procedures for the writing and speaking assessments were sufficiently reliable for screening purposes. The validity of the four assessment procedures seems sufficient, based as they are on the extant literature and, in the case of the Nelson-Denny, wide use.

Our second question dealt with how well our students measure up in each area of the assessment. The answer to this question supports the answer to question 1, as well, since discriminating between weak and strong levels of the tested skill is a requirement for content validity.

We have gathered data on a large number of variables related to this research. Table 1 presents the mean scores on the important variables. Our students score well on the national norm-referenced tests. SAT scores of 1000 are certainly respectable, as is the 60+ percentile on the
Nelson-Denny. Fewer SAT scores are reported, because some specific groups within our secondary education student population are not required to report SAT's (veterans, graduate students).

More meaningful data are seen in the comparison of students who score above and below cutoff on each of these assessments. Cutoffs were established through the first year of the assessment, using the first two cycles as the pilot group. The CRT-Reading cutoff was set at one less than maximum on each objective, or a total of 19. Writing and speaking, using the holistic scores for screening, follow the CEEB recommendations that 3 (of a possible 4) is the minimum acceptable level of competence. The 40th percentile on the Total score was adopted for the Nelson-Denny, based on the publishers' findings and the correlations between the two reading tests. The relevant correlations (reported in Table 2) are significant, even though the percent of variance accounted for is relatively low (≤9).

The percentages of students successfully completing each of the four assessments on the first try is relatively reassuring, as reported in Table 3. At least 70 percent of our prospective teachers are competent in each category. We are not surprised that 19 percent of our students succeeded in the Nelson-Denny. Such standardized tests are familiar to our students. In addition, the Nelson-Denny correlates highly (.43) with the SAT-verbal score. In that sense, the Nelson-Denny gives us relatively little information beyond that available from the SAT. The CRT-Reading test, however, causes our students more difficulty than the Nelson-Denny.
Table 1
Mean Scores of Critical Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\bar{x}$</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson-Denny Reading Test (percentiles)</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Criterion-Referenced Test - Reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>4.96</td>
<td>(6)</td>
</tr>
<tr>
<td>Comprehension - Low</td>
<td>4.74</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>7.13</td>
<td>(9)</td>
</tr>
<tr>
<td>Data Interpretation</td>
<td>2.17</td>
<td>(3)</td>
</tr>
<tr>
<td>Total</td>
<td>18.93</td>
<td>(23)</td>
</tr>
<tr>
<td>Writing Assessment</td>
<td>2.90</td>
<td>(4)</td>
</tr>
<tr>
<td>Speaking Assessment</td>
<td>2.91</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>71.77</td>
<td>(100)</td>
</tr>
<tr>
<td>SAT - Verbal</td>
<td>473</td>
<td>(142)</td>
</tr>
<tr>
<td></td>
<td>526</td>
<td></td>
</tr>
<tr>
<td></td>
<td>999</td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>2.90</td>
<td>(211)</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>CRITERION-REFERENCED TEST</th>
<th>NELSON-DENNY (%ile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>.25**</td>
</tr>
<tr>
<td>Lo Comprehension</td>
<td>-.15</td>
</tr>
<tr>
<td>Hi Comprehension</td>
<td>.30**</td>
</tr>
<tr>
<td>Data Interpretation</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.27**</td>
</tr>
</tbody>
</table>

** p < .01
Students report that it is more challenging, because of the professional nature of the reading material. The mean scores on its subtests (see Table 1) show some difficulty with both Vocabulary and High Comprehension—two reading skills we think are important in dealing with the professional literature.

Writing and speaking scores both demonstrate a high percentage of competence on this assessment. Given the rating system and the reliability levels attained consistently, it is valuable to point out the high levels of competence reported here—except for one subject area. The totals given in Table 3 would be radically different without the social studies area. Of the remaining four subject areas, 18 percent of the students are not competent in writing and 15 percent in speaking. We are, of course, pleased that students in four areas rate so highly. We are unable to account for the high percentage of prospective social studies teachers who demonstrate low levels of skill in writing and speaking.

Our research into the differences by subject area is continuing, as is our analysis of students who demonstrate deficiencies in more than one skill. To date, 63 students have registered deficiencies in one skill, 43 in two skills, 7 in all three skills (these data exclude the Nelson-Denny scores). We are also studying the retention rates of students who begin with these deficiencies.

Differences by sex. All data were treated by analysis of variance by sex, in order to determine whether the expected differences would occur. On a number of variables, the females (n=115) exceeded the males (n=107):
<table>
<thead>
<tr>
<th></th>
<th>Nelson-Denny (40th %ile)</th>
<th>CRT Reading (19)</th>
<th>CRT Writing (3)</th>
<th>Speaking (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>above</td>
<td>below</td>
<td>above</td>
<td>below</td>
</tr>
<tr>
<td>English/Communication</td>
<td>40</td>
<td>12</td>
<td>38</td>
<td>14</td>
</tr>
<tr>
<td>Math</td>
<td>31</td>
<td>9</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Science</td>
<td>45</td>
<td>7</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>Social Studies</td>
<td>49</td>
<td>17</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>10</td>
<td>1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>175</td>
<td>46</td>
<td>154</td>
<td>67</td>
</tr>
<tr>
<td>%</td>
<td>79%</td>
<td>21%</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Variable</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson-Denny Scores</td>
<td>8.512-11.893**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion-Referenced Test - Vocabulary</td>
<td>5.855*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>28.053**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>4.080*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT - Verbal</td>
<td>5.746*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>30.531**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05
** p < .01

On the SAT - Math, and all sections of the CRT except vocabulary, differences by sex were not significant.

However, in grades in relevant courses, males exceeded females in 10 of 12 courses (significant F values ranging from 6.377 to 23.313). The two courses in which difference by sex was not significant include student teaching.

This suggests that the variables used to predict success and the grades used to evaluate success respond differently to the sex variable. Both general predictors, like GPA and SAT scores, and our specific skill assessments (including criterion-referenced scores in vocabulary, writing and speaking, and the norm-referenced Nelson-Denny) favor the females in our sample. But course grades, in general, favor males. The exception of student teaching is important; since that is seen as the capstone of the teacher preparation program. We are interested in the lack of sex differences for most of the CRT-Reading test, since the traditional norm-referenced tests, like the Nelson-Denny, have consistently shown the difference reported here.
Analysis of skills assessments. One concern of this research is to establish that these assessments provide information not available from other data. Hence, we studied correlations between these basic skill assessments and the SAT scores and course grades available on our students. Table 4 reports these data. The correlations are generally significant, but at a level which suggests that much information in both variables is not explained by the correlation.

---

Insert Table 4 About Here

---

Figure 3 includes the intercorrelations of the basic skills assessments themselves. It is accepted philosophically that these three language skills are related. However, when our assessments are correlated, we do not find that strong relationship. The correlations between reading and writing range from .32 to .36, strong but not overwhelming. The correlations between writing and speaking are slightly weaker (.27-.33). The correlations between reading and speaking, all non-significant, show the weakest relationship of all. The connection between reading and speaking is tenuous, at best, given these data. However, their consistency across different tests suggests that the assumed relationships among these language skills deserve greater research.

---

Insert Figure 3 About Here

---

Conclusions

We believe that the assessment procedure described here has been demonstrated to be valid and reliable as a screening device for professional educators. The procedure has been used by different staff on five separate cycles of students. We have moved from these data to make successful
Table 4
Correlations between Basic Skills Assessments and Critical Variables
n > 113

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT - Verbal - Nelson-Denny total</td>
<td>.43**</td>
<td>.01</td>
</tr>
<tr>
<td>CRT - reading total</td>
<td>.28**</td>
<td>.05</td>
</tr>
<tr>
<td>CRT - writing</td>
<td>.30**</td>
<td>.01</td>
</tr>
<tr>
<td>Speaking</td>
<td>.08</td>
<td>.20</td>
</tr>
</tbody>
</table>

Speaking Assessment - Speech course grades

- 1-4 scale - .19
- 100-point scale - .24*

Writing Assessment - English Composition course grades

- English 10 - .21*
- English 20 - .31**

* p .05 ≥ .20
** p .01 ≥ .25
Figure 3
Correlations of Basic Skills Assessments
n > 193

READING

.32 - .36

WRITING

.27 - .33

SPEAKING

.16 - .19

p < .05 = .20
completion of this assessment prerequisite to continuing in the professional program. We have also moved to develop a second form of the criterion-referenced reading/writing test.

We believe, too, that the quality of students coming into our secondary education program is an important issue. The majority of our students can be proud of their scores, of their competence in these skills. Indeed, we make this information part of their placement record. Employing superintendents and graduate schools have indicated that they find it useful. Those students who do not meet our standards for competence are a concern. Most of them, when confronted with these results in private conferences, admit that these skills have always been a problem. Many of them transfer voluntarily to other majors. Others seek remedial help in a number of ways.

The program faculty remains convinced that students who have not demonstrated their competence should not be allowed to continue in the program. They are bolstered by the relatively high correlation between these scores and grades in student teaching.

As a result, the program faculty is undertaking a more thorough analysis of these data and other appropriate data from the student teaching experience. It is important to be able to demonstrate the connection between these basic skills assessments and success in student teaching.

Thus, in a number of ways, we are continuing this research into the basic skills of prospective teachers. We owe it to our students and their students to ensure their competence.
References


