Running head: PROJECT PARA ASSESSMENT

# Project Para Assessment
## A technical summary including passing score recommendations

Gerald Giraud, Ph.D.

Patrick Irwin, M.A.

Maggi Jo Fiske, B.S.W.

Kumar Nagarajarao, M.A.

Gabriel Giraud

July, 2004

## Purpose and test description

*Title 1 Paraprofessional non regulatory guidelines* (USDE March, 2004) state in part:

> Paraprofessionals must have the appropriate knowledge and ability to assist in instructing students and be competent in required instructional techniques and academic content areas. Paraprofessionals should also demonstrate that they are competent in basic literacy skills, including the ability to speak and write standard English. (Section C, paragraph C-1)

The Project Para Assessment is intended as a means for Nebraska local education agencies to qualify paraeducator candidates according to NCLB guidelines. The assessment consists of 6 subtests: 3 aimed at academic ability (hereafter referred to as Academic tests) in reading, writing, and mathematics, and 3 aimed at knowledge and ability related to the paraeducator role in the delivery of instruction in reading, writing and mathematics (hereafter referred to as Para Skills tests).

All test items are 4 choice multiple choice. The reading, writing and mathematics Academic tests consist of 28, 25, and 32 items respectively, and the Para Skills tests aimed at the role of paraeducators in assisting in the delivery of instruction in reading, writing, and mathematics consist of 29, 27 and 28 items respectively. Scores are determined by the number of items answered correctly, and are intended to be criterion referenced. The Academic reading and writing subtests consist of narrative passages followed by related multiple choice items. All other subtests consist of stand alone items.

Plans are for the Project Para Assessment to be administered in an online format in which examinees make and submit answer choices via computer. The online administration of the assessment will not be managed by the authors of this report. The assessment will be proctored by local education agency personnel, and scoring will be automatic.

## Test development

### *Academic subtests*

Test content specifications for the academic reading, writing and mathematics tests were derived from the Nebraska 12$^{th}$ grade content standards (www.nde.state.us/Acadstand.html). The standards were modified to reflect a cognitive level equivalent to two years of post-secondary education pursuant to *Title 1 Paraprofessional non regulatory guidelines* (USDE March, 2004).

Items for the academic reading and writing tests were written by an experienced test writer with graduate level education in reading and writing instruction and assessment. The item writer based the assessment construction on the above referenced content standards, aiming the assessment text and items at the cognitive level expected of two year postsecondary level students. Items for the academic mathematics assessment were written by a graduate level prepared mathematics instructor with experience in teaching and assessment development at the secondary and post secondary level.

Items for all academic subtests were preliminarily reviewed by experienced Nebraskan editors for form and grammar. Complete draft subtests were then reviewed for match to test specifications and biased content by Ph.D. degreed postsecondary educators in reading, writing and mathematics experienced with Nebraska students at the two year college level. These reviewers were not involved in item writing. Before reviewing the items for bias, reviewers were provided with written guidelines for bias review, and completed a quiz related to bias review (see Appendix A). Some items were edited or modified based on the results of the match to test specification review.

*Para Skills subtests*

In August 2003, a panel of experienced Nebraska paraeducators and teachers familiar with the role of paraeducators in the classroom (hereafter referred to as the Para Skills panel) collaboratively developed sets of knowledge and skills relevant to the role of paraeducators in assisting in the instruction of reading, writing and mathematics. The knowledge and skill sets developed by the Para Skills panel became the basis for the development of both Project Para training materials and Project Para Assessment related to the paraeducator role in assisting instruction in reading, writing and mathematics. Information concerning the development of the training materials is not part of this report, but it should be noted that the training materials were used as a basis for the development of subtest items.

Items aimed at the knowledge and skills identified by the Para Skills panel were written by Nebraska educators with graduate level training in instruction and curriculum in the content areas of reading, writing and mathematics and who are familiar with the role of paraeducators in the classroom. Before writing items, item writers reviewed item writing instructional materials, as well as materials related to bias in item construction (see Appendix A), and completed a quiz on item construction and item bias. The item writers used the Project Para training materials that had been developed based on the knowledge and skills identified by the Para Skills panel as a resource for writing items. Items for all Para Skills subtests were preliminarily reviewed by experienced editors for form and grammar.

Complete draft subtests were next reviewed for match to test specification and bias by a panel of 19 experienced Nebraska paraeducators who had not been involved in the test development process. Before reviewing items for match to test specification, reviewers discussed the test specifications and reached consensus on example items'

match to specifications for the Para Skills reading, writing and mathematics subtests. Item match to specification was determined by 10 of 19 reviewers agreeing in independent judgments that the skills, knowledge and cognitive processes measured by an item matched the relevant specification. Results of match to specification review are reported in Tables 1-3. The number of items shown to match specifications in these tables refers to items used in the final drafts of Para Skills subtests.

Before reviewing the items for bias, reviewers were provided with training in bias identification and guidelines for bias review, and completed a quiz related to the bias review process (see Appendix B). Some items were edited or modified based on the results of the match to specification and bias review.

## Pilot administration

Pilot test forms for Academic and Para Skills subtests were constructed through processes described above. The pilot draft Academic subtests for reading, writing and mathematics consisted of 30, 25 and 34 items respectively. The pilot draft Para Skills subtests for assisting instruction in reading, writing and mathematics consisted of 39, 30 and 39 items respectively.

Pilot draft subtest forms and scanner scored answer sheets were distributed by mail to 300 Nebraskan examinees. Of these, 232 returned usable answer sets, for a return rate of 77.6 %. Occupation and years of school for pilot examinees who returned usable answer sets are reported in Table 4. Ethnicity by occupation is reported in Table 5.

Results of item and reliability analyses of data from the pilot administration of the Project Para Assessment were used to create final draft forms of the Academic and Para Skills subtests. All item and total score analyses were conducted using methods and decision algorithms consistent with classical test theory.

## Final draft subtest item and reliability technical information

### *Academic Reading subtest*

The final draft of the Academic Reading subtest is comprised of 28 items. Two items were discarded from the pilot form because of poor performance.

Item discrimination analysis based on pilot data reveals that for all items the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performers (see Table 6). Mean item difficulty was .811 (SD = .14).

Mean corrected item-total correlation was .32, with one retained item having a corrected item-total correlation of .10. The highest corrected item-total correlation was .51. Reliability as indicated by standardized item alpha is .8129.

*Academic Writing subtest*

The final draft of the Academic Writing subtest is comprised of 25 items. All items were retained from the pilot form.

Item discrimination analysis based on pilot data reveals that for all items save one the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performing (see Table 7). Item performance on one item was equivalent. Mean item difficulty was .791 (SD = .18). Mean corrected item-total correlation was .29, with one retained item having a corrected item-total correlation of .13. The highest corrected item-total correlation was .46. Reliability as indicated by standardized item alpha is .77.

*Academic Mathematics subtest*

The final draft of the Academic Mathematics subtest is comprised of 32 items. Two items were discarded from the pilot form because of poor performance.

Item analysis based on pilot data reveals that for all items the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performing (see Table 8). Mean item difficulty was .77 (SD = .14).

Mean corrected item-total correlation was .37, with one retained item having a corrected item-total correlation of .12. The highest corrected item-total correlation was .58. Reliability as indicated by standardized item alpha is .86.

*Para Skills Reading subtest*

The final draft of the Para Skills Reading subtest is comprised of 28 items. Eleven items were discarded from the pilot form because of poor performance.

Item analysis based on pilot data reveals that for all items the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performing (see Table 9). Mean item difficulty was .72 (SD = .18).

Mean corrected item-total correlation was .32, with one retained item having a corrected item-total correlation of .21. The highest corrected item-total correlation was .48. Reliability as indicated by standardized item alpha is .81.

*Para Skills Writing subtest*

The final draft of the Para Skills Writing subtest is comprised of 27 items. Three items were discarded from the pilot form because of poor performance.

Item analysis based on pilot data reveals that for all items the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performing (see Table 10). Mean item difficulty was .67 (SD = .20).

Mean corrected item-total correlation was .27, with one retained item having a corrected item-total correlation of .14. The highest corrected item-total correlation was .41. Reliability as indicated by standardized item alpha is .75.

*Para Skills Mathematics subtest*

The final draft of the Para Skills Mathematics subtest is comprised of 28 items. Eleven items were discarded from the pilot form because of poor performance.

Item analysis based on pilot data reveals that for all items the performance of the top performing pilot examinees (as determined by total subtest score) was significantly better than the performance of the lowest performing (see Table 11). Mean item difficulty was .82 (SD = .15).

Mean corrected item-total correlation was .31, with one retained item having a corrected item-total correlation of .11. The highest corrected item-total correlation was .54. Reliability as indicated by standardized item alpha is .80.

*Combined score reliability*

Because passing scores for Project Para Assessments might be based on combined subtest scores, reliability of combined scores is of interest. Three combined score reliabilities are reported here: Academic subtest combined score (standardized item alpha = .9130), Para Skills subtest combined score (standardized item alpha = .9116), and Project Para total score (standardized item alpha = .9278)

.

## Test score validity

*Content validity*

Because Project Para Assessment is structured as an objective test of knowledge and skills, content validity is of interest. Evidence for content validity was gathered by independent review of test items for match to test specifications and biased content by paraeducator and academic content experts as described above.

*Constructl validity*

Evidence to support the contention that subtest scores measured unique constructs was obtained through examination of relationships among subtests scores. Correlations among subtest scores are reported in Table 12. Correlations among subtest scores were moderate, suggesting the presence of unique variance in subtest scores. Evidence that

Para Skills subtests and Academic subtests measure respectively unique constructs is observed in the correlation matrix presented in Table 12. Note the correlations among Para Skills are higher than correlations between Para Skills subtests and Academic subtests.

Factor analysis utilizing direct oblimin rotation (on the assumption that extracted factors are expected to be correlated) reveals two factors, with Academic subtests loading on one factor and Para Skills subtests loading on the other (see Table 13). Factors account for 75 % of total variance (10% and 65 % respectively). Results suggest a strong Para Skills factor accompanied by an Academic factor.

If the Project Para Assessment scores do aid in the identification of qualified paraeducators, then it is logical to expect that teacher performance on the subtests will exceed the performance of other examinees, and that paraeducator performance will exceed the performance of non teachers with similar educational background. Means and standard deviations of subtest scores by examinee classification in terms of educational role (teacher, paraeducator, other) and years of education are reported in Tables 14 (Academic subtests) and 15 (Para Skills subtests). Results for total Academic and Para Skills and total Project Para Assessment scores are reported in Table 16.

*Criterion validity*

By criterion validity, we mean that the decisions about examinees made based on Project Para Assessment scores are defensible. In the present case, scores are intended to identify examinees as qualified to perform as paraeducator. An examination of the fail rates that would accrue based on defensible passing scores presented below reveals that only 1 of 21 teachers who participated in the pilot administration would fail any subtest, even at the highest defensible passing scores. This supports the validity of basing judgments about the qualifications of examinees relative to education on the assessment scores.

## Passing Scores

Project Para Assessment score interpretation is intended to be criterion referenced. Thus, passing score options were determined for all 6 subtests. Passing score options were determined through a modified Angoff (1971) method, a test centered pass score setting method. Experienced paraeducators (n=16) participated in a facilitated workshop in which judgments concerning the performance of the qualified paraeducator candidate were elicited and recorded. The workshop process was facilitated by a Ph.D. psychometrician assisted by an experienced M.A. psychometrician. Through a planned process modeled on peer reviewed best practice, the paraeducator panelists examined each subtest item by item and made a judgment about the likely performance of the qualified paraeducator candidate taking into account the characteristics of the qualified paraeducator and the difficulty of the item. The passing scores suggested here are the averages of individual panelists' passing scores. It is important to note that final official passing scores are to be determined by appropriate policy making entities.

Passing score options for policy makers are several. Reported below for each subtest are the mean of panelists passing scores derived from the processes described above, the median, and values 1 and 2 standard errors above and below the mean panelist passing score all expressed in nearest whole values. For each subtest, the percent of pilot test examinees and the percent of pilot examinees self identified as currently working paraeducators that would fail, given the passing scores, is reported. In addition, passing scores for combined subtests are presented with passing scores for the total combined set of subtests.

*Academic Reading subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 18 | 8.6 | 9.1 |
| Mean -1 SE | 19 | 12.1 | 14.5 |
| Mean | 20 (SE = 1) | 16.4 | 16.4 |
| Mean + 1 SE | 21 | 24.6 | 26.4 |
| Mean + 2 SE | 22 | 31.5 | 36.4 |
| Median | 18.5 | see 18 or 19 above | |

*Academic Writing subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 17 | 12.9 | 10.0 |
| Mean -1 SE | 18 | 19.3 | 17.3 |
| Mean | 19 (SE = 1) | 30.9 | 30.9 |
| Mean + 1 SE | 20 | 40.3 | 41.8 |
| Mean + 2 SE | 21 | 54.1 | 58.2 |
| Median | 18 | see 18 above | |

*Academic Mathematics subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 20 | 16.3 | 15.5 |
| Mean -1 SE | 21 | 20.2 | 21.8 |
| Mean | 22 (SE = 1) | 22.7 | 24.5 |
| Mean + 1 SE | 23 | 28.3 | 29.1 |
| Mean + 2 SE | 24 | 33.9 | 33.6 |
| Median | 20.5 | see 21 or 20 above | |

*Para Skills Reading subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 16 | 13.3 | 10.9 |
| Mean -1 SE | 17 | 17.6 | 13.6 |
| Mean | 18 (SE = 1) | 22.7 | 20.9 |
| Mean + 1 SE | 19 | 29.2 | 29.1 |
| Mean + 2 SE | 20 | 35.2 | 38.2 |
| Median | 16.5 | see 16 or 17 above | |

*Para Skills Writing subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 13 | 9.1 | 10.0 |
| Mean -1 SE | 14 | 13.4 | 16.4 |
| Mean | 15 (SE = 1) | 19.8 | 23.6 |
| Mean + 1 SE | 16 | 26.7 | 31.8 |
| Mean + 2 SE | 17 | 35.8 | 40.9 |
| Median | 14.5 | see 14 or 15 above | |

*Para Skills Mathematics subtest*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 20 | 11.6 | 10.9 |
| Mean -1 SE | 21 | 17.6 | 20.0 |
| Mean | 22 (SE = 1) | 24.9 | 25.5 |
| Mean + 1 SE | 23 | 35.2 | 37.3 |
| Mean + 2 SE | 24 | 46.4 | 47.3 |
| Median | 21.5 | see 21 or 22 above | |

*Academic total score*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 55 | 12.9 | 13.6 |
| Mean -1 SE | 58 | 16.7 | 17.3 |
| Mean | 61 (SE = 3) | 22.3 | 20.0 |
| Mean + 1 SE | 64 | 35.6 | 28.2 |
| Mean + 2 SE | 67 | 41.2 | 44.5 |
| Median | 60 | 18.9 | 19.1 |

*Para Skills total score*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 49 | 11.2 | 8.2 |
| Mean -1 SE | 52 | 14.6 | 11.8 |
| Mean | 55 (SE = 3) | 19.7 | 18.2 |
| Mean + 1 SE | 58 | 31.8 | 33.6 |
| Mean + 2 SE | 61 | 42.1 | 46.4 |
| Median | 53 | 18.5 | 15.5 |

*Project Para Assessment Total score*

| | Passing score | % failing all examinees | % failing para examinees |
|---|---|---|---|
| Mean -2 SE | 110 | 16.3 | 15.5 |
| Mean -1 SE | 113 | 18.9 | 18.2 |
| Mean | 116 (SE = 3) | 21.0 | 20.9 |
| Mean + 1 SE | 119 | 24.9 | 24.5 |
| Mean + 2 SE | 122 | 29.6 | 32.3 |
| Median | 116.5 | 21.0 | 20.9 |

## Passing score options

The passing score for Project Para assessment can be set in different ways. Psychometrically defensible passing scores for six subtests are suggested. Passing scores for combined Academic test score and for combined Para Skills score are also suggested. Finally, passing score options are presented for the combined total Project Para Assessment score. Among commonly considered factors in determining passing scores are the percent of likely examinees who might fail at different defensible passing scores, and the impact that might have on local education agencies. One option might be to set one passing score for the combined Academic tests, and then a separate passing score for each of the 3 Para Skills subtests, on the theory that candidates may seek training to improve their knowledge and skills in a particular subtest content domain having passed one or two subtests but failed others.

The authors of this report recommend that defensible passing scores at or below the mean or median scores presented here be seriously considered. For example, the passing score suggested by the mean pass score derived from the Angoff (1971) described above is 116. Empirical evidence to support this level of passing score can be found in the comparison of mean test scores for Project Para Assessment total score in Table 16. The mean total score for the lowest educational level of currently practicing paraeducators was approximately 122, compared to 109 for non educator examinees at the same education level. Thus, a passing score of 116 would fail fewer paraeducators than non paraeducators at the same educational level. For example, at the education level of 12-13 years, a pass score of 116 would fail 60 % of examinees in the non educator category, and only 30% of examinees who are currently paraeducators. At the 14 year education level (the target of the Academic subtests), 30 % of non educator examinees fail at a pass score of 116, and only 15% of paraeducators. Refer to the Project Para Assessment total score frequency distribution in Appendix C for other comparisons.

## Conclusion and other recommendations

Project Para Assessment results in reliable and arguably valid scores for the purpose of identifying qualified paraeducator candidates. While some subtest scores have marginal reliability (.75 to .86), combined scores exceed .90. It should be remembered that Project Para Assessment scores are met to be criterion referenced. Thus, as Project Para Assessment is used in practice, criterion referenced strategies for determining reliability should be pursued.

The authors of this report recommend that item development and test evaluation continue as Project Para is put into operation. Ideally, a bank of items will be developed so that the advantages of online administration can be realized, including the ability to deliver randomly selected equivalent test forms. This will require continued attention to test score data and a commitment to item development and piloting. Consideration should be given to the development of assessments that are aimed at the different grade levels in which paraeducators work.

The authors further recommend that performance tasks be incorporated into Project Para Assessment. The development of valid and reliable performance assessment will require a substantial commitment of time and resources, but would result in a more authentic and informative evaluation of pareducator candidates.

Finally, it is important to note that a score derived from this assessment, or any assessment result, is only one piece of evidence in decision making. Exceeding the pass score on this or any assessment should not be the only factor in hiring. Interviews, references and observation of performance are also essential in identifying those who are suited for the work of a paraeducator.