

Project Para Assessment Content Validity and Bias Review Report
Para Skills Subtests

Gerald Giraud, Ph.D.

July 30, 2004

Introduction

The purpose of this report is to document the procedures and analyses undertaken to determine the content validity and presence of bias in the Project Para Assessment. The report summarizes the procedures and the results of this validity and bias study.

Purpose and test description

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 (reading = 40 items, writing = 30 items, mathematics = 40 items) 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. Panelists were recruited from representative geographic regions of the state. Panelist practiced as paraeducators in a range of grade levels from elementary to high school. The panelists had a range of experience of from 1 to 20 years as a paraeducator, with an average of 8 years experience. Eleven panelists had at least some college, and 3 had 4 year college degrees.

Methods and Procedures

The workshop, held in Kearney, Nebraska, on June 3, 2004, included in order the following activities: orientation, training in bias review, bias review activity, training in matching items to test specification, item match to specification activity, conclusion and workshop evaluation.

The workshop began with Gerald Giraud (workshop facilitator) discussing the importance of the content validity task and the procedures that would take place over the next few hours. Dr. Giraud presented an overview of Project Para Assessment and provided a general description of the bias and content review process.

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 A). The bias review quiz and practice items were discussed by panelists in small groups, and then by the entire panel facilitated by Dr. Giraud. Panelists reached consensus on the quiz and practice items.

Following training in bias identification, panelists reviewed Para Skills subtest drafts item by item and made independent judgments about the presence of biased content or language. Judgments were made independently by panelists, without discussion with other panelists. Panelists were also asked to rate the subtests overall in regard to biased content.

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. The content, skills and cognitive processes involved in each test specification were discussed, and items were examined to determine the extent to which the skills, knowledge and cognitive process called for in each practice item matched a test specification. Practice items were discussed in small groups, and then the entire panel, led by Dr. Giraud, discussed item match to specification. Group consensus was reached on the match of practice items to subtest specifications.

Following training, copies of the draft Para Skills reading subtest were distributed to panelists, along with forms listing the reading subtest specifications and item numbers. Panelists were instructed to indicate item to specification match by marking an H (high match), M (medium match) or L (low match) in the appropriate item/specification cell. Panelists were told that items could match more than one specification. Panelist made the item to specification match independently without discussion with other panelists. Each Para Skills subtest (reading, writing and mathematics) was reviewed for item to specification match in turn.

At the conclusion of workshop, panelist completed an evaluation of the workshop activities. All panelists received certificates of participation, and all workshop materials were collected by the workshop facilitator.

Results

Bias review

Biased content was identified in one mathematics item, in which a student named Jorge incorrectly answered a math question. The reference to Jorge was changed to

generic reference to a student. No other biased was identified by any panelist. Some panelists did find typographical or grammatical errors in some items, and these were recorded and subsequently corrected. All panelists reported that as a whole, the subtests were not biased toward any likely group of examinees.

Item match to specification

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 (either high match or moderate match) 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.

Workshop evaluation

At the conclusion of the Passing Score Setting Workshop, panelists completed an evaluation form consisting of 3 parts. Part 1 focused on the orientation and training; Part 2 focused on the judgments made by panelists and on the levels of confidence in making the bias and match to specification judgments and on the amount of time allowed to make the judgments. The overall workshop quality was assessed in Part 3. Results for the workshop evaluation are reported in Table 4.

Part 1: Training. Training was evaluated on a scale ranging from 1 - 6, where 1 = Very Unsuccessful and 6 = Very Successful. On average, the panelists rated all components of the training greater than 4.5 .

Panelists also rated the adequacy of the time provided for training and orientation. on a six-point scale, where 1 = Totally Inadequate and 6 = Totally Adequate. All ratings exceeded 4.5. (See Table 4). Panelists also rated the overall amount of time devoted to training and orientation on a 3 point scale, with 1= not enough time, and 3 = too much time. The average rating was approximately 2 (the right amount of time).

Part 2: Judgments. Panelists were asked to rate the adequacy of time allocated for reviewing items for bias and match to content on a four-point scale (1 = More time needed and 4 = More than enough time was allotted), the average rating was for bias review was 3.1, and for item to specification match 2.95.

Panelists were asked to rate confidence in their ability to provide meaningful judgments for bias and item match to specification on a four-point scale (1 = Not Confident and 4 = Confident). Average ratings were 3.1 for bias review, and 2.65 for item to specification match. The average panelist was at least somewhat confident in the judgments made about bias, and somewhat less confident about item to specification match (a more difficult cognitive task).

Part 3: Overall Rating. Two questions asked panelists to rate (1 = Totally Unsuccessful and 4 = Totally Successful) the success and organization of the workshop. The average ratings on these two items was 2.86 and 3.13, respectively.

In sum, the panelists felt on average that the training was adequate, the workshop was successful and that they were moderately confident in the judgments made during the workshop.

Conclusion

The results of this workshop provide evidence that the Para Skills test items do not contain biased content and do match the specifications for the Para Skills assessment. Each specification save one is matched by at least two items. It is important to note that for one math specification (i.e. Explain the levels of instruction in mathematics learning (see Table 3)), no item match was identified. This underlines the importance of continued item development, and the need for inclusion of performance assessment as a supplement to these objective subtests. However, the Para Skills subtests are intended to be tests of general knowledge and skills related to the paraeducator role in assisting in the instruction of reading, writing and mathematics, and the results of this study support this interpretation of test results. It is not appropriate to make decisions about examinees at the level of specification, as the tests have too few items to make such inferences.

Table 1. Number of items determined to match Para Skills reading subtest.

Para Skills Reading subtest specifications	# of final form items match (items can match more than one item)
Describe key developments in reading instruction	3
Describe important terms used in reading instruction	13
Demonstrate skills in reading aloud to students	2
Demonstrate strategies to help children read unfamiliar words	4
Help students read smoothly without pauses and hesitation	3
Identify methods for helping students understand what they read	5

Table 2. Number of items determined to match Para Skills writing subtest.

Para Skills Writing subtest specifications	# of final form items match (items can match more than one item)
Identify the components of a trait-based writing program	9
Discuss the development of written language skills including spelling and handwriting and their relationship and importance in written expression	12
Describe how a paraeducator might use strategies, assistive devices, and aids to reinforce teacher instruction in written expression	8

Table 3. Number of items determined to match Para Skills math subtest

Para Skills Math subtest specifications	# of final form items match (items can match more than one item)
State the role of the teacher and para-educator in facilitating mathematics development.	3
Describe the stages of mathematics development	2
Demonstrate a basic understanding of frequently used mathematics terms, concepts, and symbols	10
Explain the levels of instruction in mathematics learning	
Demonstrate the ability to reinforce the math instruction presented by the teacher in readiness, basic computation and problem solving skills, as well as in advanced mathematical skills	7
Demonstrate the ability to analyze, provide feedback, and assist in correcting student math errors	3
Explain appropriate use of math aids such as calculators	4

Table 4. Workshop evaluation summary results.

	Minimum	Maximum	Mean
Orientation Training Success	4.00	6.00	5.3043
Bias Review Training Success	4.00	6.00	5.4348
Validity Training Success	3.00	6.00	4.9565
Validity Practice Success	2.00	6.00	4.5217
Overall Training Success	1.00	6.00	4.7391
Orientation time	3.00	6.00	5.5217
Bias Review Training Time	5.00	6.00	5.6522
Validity Training Time	3.00	6.00	5.2174
Validity Practice Time	2.00	6.00	4.8696
Overall Training time	1.00	6.00	5.1304
Rate the amount of time	2.00	3.00	2.0870
Time allotted to bias review	3.00	4.00	3.1304
Confidence in bias review	3.00	4.00	3.6522
Time allotted to item to objective match	1.00	4.00	2.9565
Confidence in item to objective match	1.00	4.00	2.6522
Overall rating of workshop	2.00	3.00	2.8696
Rate the organization of the workshop	2.00	4.00	3.1304

Appendix A

Materials from Para Skills content validity and bias review

Introduction to item review workshop

The purpose of item review is multi-faceted. Items are reviewed for biased content, appropriate content, relevance and structure. The goal is to end up with a pool of items that effectively and fairly evaluate test takers' knowledge of content essential to practice as a Para.

As a participant in item review, you will read materials that inform you of criteria for each aspect of item review, and participate in online discussions related to these materials. You will then review items and record your ratings on forms provided for this purpose.

Bias Review

When knowledge is measured using multiple choice or other item formats, it is assumed that these items measure relevant aspects of the test takers' knowledge, and not other, irrelevant things, like emotional reaction to item wording or content. Further, it is assumed that anyone with the knowledge intended to be measured has the same opportunity to answer correctly, and that some test takers will not be advantaged or disadvantaged by having particular characteristics, such as gender or ethnicity, that are not relevant to the content being assessed. You will review items for biased content. Any items that are judged to exhibit bias will be revised or deleted from the final version of the test.

Structure review

It is important that items be correct in terms of grammar, spelling and structure. This part of your review involves flagging items for grammar, spelling and structure problems as you review them.

Relevance review

Items are intended to assess Para's understanding of important objectives related to practicing as a Para in the classroom. You will review items and match items to the objective you believe is most closely measured by the item. You will also rate the relevance and importance of each item to the objective that it measures.

Identifying Bias in Test items

Project Para

What is Bias in test items?

Bias is the presence of some characteristic of an item that results in differential performance for individuals of the same ability but from different ethnic, sex, cultural, or religious groups¹

In general, bias comes in many forms. It can be age, gender, race, ethnicity, cultural or disabilities and mental condition. In testing scenarios, an item may be biased if it contains content or language that is differentially familiar to subgroups of examinees. Therefore, item bias is divided into two types content bias and language bias.

An example of content bias against girls would be one in which students are asked to compare the weights of several objects, including a football. Since girls are less likely to have handled a football, they might find the item more difficult than boys, even though they have mastered the concept measured by the item (Scheuneman, 1982a).

An item may be language biased if it uses terms that are not commonly used statewide or if it uses terms that have different connotations in different parts of the state. The language used for items should have the same basic semantic content for all persons regardless of age, race, gender, ethnicity or physical or mental condition. Item reviewers should consider the connotation of words, expressions, and symbols and avoid negative and potentially inflammatory connotations. For example, terms that are generally unacceptable in test items include lower class, mentally retarded, housewife, colored people, and so on. Another example of language bias: items that include reference to silos or corn rows could have different meaning depending on whether the reader has a rural or urban background.

Guidelines for identifying bias in test items

In this section the guidelines are grouped into stereotyping, gender, ethnicity, disabilities and mental condition, and age.

Stereotyping

Test materials *should not convey* any of the following.

- That a population group is deserving of a particular fate.
- That a population group is by nature dependent on help from the majority culture.
- That a population group lacks some positive quality fairly common to humans.
- That a population group has excess of a quality fairly common to humans.
- That a population group has a genetic deficiency or surplus in some area of intellect, talent or ability.
- That particular groups of people are by nature poor, criminal, more or less intelligent, good dancers, adept at certain tasks, etc.

Gender

- *Use gender-neutral language.* Example, use police officer and flight attendant as opposed to policeman and stewardess.
- In traditional usage the term *man* means both '*adult male human being*' and '*human being*'. The generic 'man' can foster ambiguities therefore it is better to use terms such as humanity, mankind, human, people and so on.
- *Avoid compound forms employing 'man'* - for example, manpower, manmade, craftsman, fireman, and so on when they refer to both sexes. Employ gender-neutral alternatives such as personnel, artisan, manufactured, fire fighter and so on.
- Sometimes readers interpret the word *he* as male even when it is being used in the generic sense. A double pronoun such as *he or she, he/she, s/he* can be used instead of *he* to avoid confusion.

- Avoid implications that some interests and activities are restricted to or favored by males or females.
- Avoid implications that holding a particular job is incompatible with being a man or a woman.
- Portray both males and females as persons who pursue interests based on their individuality rather than their gender.

Disabilities and mental condition

- Avoid terms that have negative connotations such as crippled, deficient, defective, deformed, lame, maimed and unfit.
- Avoid disabilities that serve as metaphors in phrases such as *deaf to our wishes* or *blind to the truth*.
- Avoid portraying persons with disabilities as objects of curiosity.

Race, Ethnicity and Religion

- Make no suggestion that any racial, ethnic, or religious group is more or less worthy than any other.
- Be sensitive to what names minority or cultural groups prefer.
- Balance depictions of people in what are traditional roles with depictions of people in nontraditional roles.
- Convey wherever possible the differences between diverse minority groups are valuable.
- Avoid derogatory regional designations (e.g. hick, hillbilly, redneck).
- Make no attempt to play down or hide history; materials should be presented accurately and not evaluated with values of this present day.

Age

- According to most guidelines, *boy and girl* apply to individuals up to about the age of thirteen or fourteen.
- *Youth* may apply both to a person of either sex between the ages of thirteen and nineteen.

- It is preferable to use the term *elder* and *elderly person* instead of *older person* in referring to an individual past the middle age.

IDENTIFY THE TYPE OF BIAS IN THE FOLLOWING EXAMPLES.

Question1:

A test is gender biased if men and women with the _____ ability levels tend to obtain _____ scores.

- a. Different, same
- b. Same, different
- c. Different, different
- d. Same, same

Question2:

If test questions on topics such as sports and business appeared more than questions on topics such as relationships and humanities, then the test might be

- a. Biased for race.
- b. Biased for gender.
- c. Biased for ethnicity.
- d. Biased for mental condition.

Question3:

In the sentence 'A nurse must undergo rigorous training before she is licensed', is the nurse male or female?

- Male
- Female

How could the question be changed to be gender neutral?

Question4:

In the following sentence, fill in the missing words from the choices given below.

A test item can be considered to be _____ if all individuals having _____ underlying ability level have an _____ probability of correctly answering the item, regardless of their subgroup membership.

- Biased, the same, equal
- Unbiased, the same, equal
- Unbiased, the same, unequal
- Biased, different, equal.

Question5:

Using bias-free language guidelines, complete the sequence for the following.

- Draftsman: Drafter:: Chairman: _____
- Layman: Layperson:: Watchman: _____

Question6:

According to most guidelines, boy and girl apply to individuals up to about the age of _____

- 15 or 17
- 18 and above
- 13 or 14
- 15 or 16

Question7:

The statement *'each student should hand in his term paper by next month'* indicates which kind of bias.

- Language
- Gender
- Disabilities and mental condition
- Age

The statement can be rephrased into: _____

Question8:

The statement *'Mrs. Jones told Jose to mow the lawn'* indicates which kind of bias.

- Race
- Gender
- Age
- Stereotyping

The statement can be rephrased into: _____

Question9:

In the statement *'Not surprisingly, Asian-American students did best in the math contest'*, which word is indicating bias?

- Asian-American
- Math contest
- Not surprisingly
- Did best

Question10:

Which among the following is the preferred-way for stating *'the visually impaired student used a special keyboard'*?

- The student, who is visually impaired used a special keyboard.
- The blind student used a special keyboard.

Question11:

Which among the following is the preferred-way for stating *'She's a good basketball player. She shoots like a man'*?

- She's a good basketball player. Like a man she shoots well.
- She's a good basketball player because she shoots well like a man.
- She shoots well like a man. She's a good basketball player.
- She's a good basketball player. She shoots well.

Question12

The farmer filled his forage pit with milo, and baled his alfalfa into large round bales.
What is the potential problem with this item?