SUBCOMMITTEE ON DISABILITY ASSISTANCE AND MEMORIAL AFFAIRS
September 13, 2006

VBA SKILLS CERTIFICATION TESTING PROGRAM
Lauress L. Wise, PhD.
President and CEO of The Human Resources Research Organization (HumRRO)

Written Testimony

Good afternoon. I am Lauress L. Wise, the president and CEO of the Human Resources Research Organization, known less formally as HumRRO. HumRRO is a non-profit, 501(c)3 research and development organization, established in 1951, that works with government agencies and other organizations to improve their effectiveness through improved human capital development and management.

I have been asked to testify today about work that HumRRO has done for the Veterans Benefit Administration (VBA) on their program for certifying essential skills for Veterans Service Representatives. These service representatives play a key role in seeing that our veterans receive the full array of benefits to which they are entitled. VSR performance at the highest level of the position requires a thorough understanding of an extensive set of policies and procedures concerning veterans' benefits and skill in identifying appropriate applications of these procedures to individual circumstances. The skills certification program embarked on by the VBA is critical to ensuring that service representatives have the knowledge and skills needed to perform their jobs effectively.

Development of the VSR Skills Certification Test

In January 2001, the VBA contracted with HumRRO to assist in the design, development, and validation of an effective and defensible certification process for the VSR position. HumRRO has worked with VBA to develop a certification program that assesses the knowledge of GS-996-10 incumbents to judge their readiness for promotion to the GS-11 position. GS-10 VSRs who pass the certification test are promoted to the GS-11 position; GS-11s who pass the test receive a bonus.

Job Analysis

During 2001, HumRRO conducted an extensive analysis of the VSR job. We worked with senior incumbents to identify critical VSR tasks, rate their importance, and identify the knowledge and skills needed to perform these tasks effectively. The critical tasks were organized into functional areas identified as important by the VBA Design Team[1]. These areas included: (a) Compensation, (b) Pension, (c) Public Contact, (d) Administrative Decisions, and (5)Appeals.

Development of Test Questions

The Design Team used the results of the survey to develop a test blueprint, which specified the number of test questions needed to cover each of the functional areas. We then trained the Design Team to write high quality test questions (items) and conducted several item development workshops to review and revise these questions. HumRRO worked with the VBA to conduct a pilot test of the test questions and test administration procedures. Many questions were dropped after the pilot test either because the item statistics were less than optimal or because pilot test participants indicated problems with a question. This is the norm; we typically develop about three times the number of items we need for administration, knowing from experience that we will lose over half in revision or piloting.

Changes to the VSR Job

When the Claims Processing Task Force Report was published in the Fall of 2001, the certification program was put on hold while recommendations from the report were put into place. The Claims Process Improvement (CPI) initiative that followed included some significant changes to the VSR position. In April 2002, VBA contracted with HumRRO to conduct several site visits to determine whether the test items, which had been written at a time when the VSR job was a generalist position, were still appropriate for VSRs who were now working on specialized teams.

Following the site visits, HumRRO met with representatives of VBA, Compensation and Pension (C&P) training, and the VBA Central Office to discuss the impact of CPI on training and skills certification. The decision was made to proceed with the generalist test because the

---

[1] The VBA Design Team represented all major stakeholders in the claims processing field (i.e., VBA management, AFGE, the Compensation and Pension line of business, incumbents, and veterans service organizations)

policy was that VSRs would be rotated across teams to maintain the skills acquired in training. A GS-11 in this position can be assigned to any team based on the needs of the station and small stations may only have one or two GS-11s. These GS-11s must be capable of reviewing and authorizing all of the work performed at the station regardless of the team from which it originated. Specialized tests reflecting specific team assignments would not tap skills that would be needed for future assignments, so HumRRO recommended that work continue using the general test blueprints previously established.

Restarting the Program

In the Fall of 2002, VBA put together a new Design Team whose task it was to get the certification process moving again. The Design Team reviewed the test blueprint, the Candidate Guide, Test Administrator Manual, and other test support documents (e.g., background information forms, confidentiality agreements). These support documents were updated to reflect changes in the program in the intervening years. The Design Team also reviewed the test questions and dropped some due to changes in the VSR job. They also wrote new questions to take the place of those that were dropped. These new items were pilot tested in February, 2003 in preparation for a spring test. This pilot test used the updated support documents, which would also be used in that test.

Operational Field Test

An operational field test was conducted in August 2003 that involved administering an over-length version of the skills certification test to 298 eligible GS-10 and GS-11 VSRs. The operational exam is designed to include 100 operational questions; we administered two over-length exams (about 120 items each) to allow us to collect data on all the items in the item bank so they would be ready for use in future administrations. HumRRO staff identified a set of 100 questions for each of the two forms that met the test specifications and demonstrated solid statistical properties, and computed overall scores based on the selected items.

Passing Score

After the field test was completed, subject matter experts (senior VBA employees who had been promoted from the GS-11 position and were "Super Senior" VSRs or Ratings VSRs)

participated in a workshop to establish a minimum passing score for the test. HumRRO used an established standard setting procedure that required experts to estimate, for each question, the percent of minimally qualified examinees who would answer the question correctly.

To pass the certification, candidates had to pass the whole test, as well as meet minimum score requirements on the compensation and pension subtests. Based on the standard setting results, candidates were required to correctly answer about three-quarters of all of the questions, three-quarters of the compensation questions, and just over half of the pension questions to pass the exam. Seventy-five candidates (25%) passed all three hurdles. While VBA had hoped for a higher pass rate, they verified with management at several Regional Offices that candidates who passed were those who were expected to do so, and those who failed were expected to have difficulty meeting the certification requirements.

Subsequent to the field test, test blueprints were revised giving more emphasis to compensation and less to pensions. Another standard setting workshop was held to establish minimum passing scores for the first operational administration in May 2006. Based on results from this workshop, candidates were required to correctly answer two-thirds of all questions and also two-thirds of the compensation questions to pass the test. The separate requirement based on pension questions by themselves was dropped.

Criterion-Related Validation Study

In 2004, the Office of Personnel Management reviewed the VSR Skills Certification Program to determine whether there were potential problems with using it as part of the promotion process. The overall passing rate in the field test was generally low, about 25%. A particular concern was that the passing rate for African Americans was significantly lower than for other incumbents. When a test results in this type of adverse impact for a particular group, legal guidelines require employers to demonstrate that test scores are a valid reflection of the skills needed to perform the job. While HumRRO had previously collected content validity data showing the relevance of each of the test questions, the VBA decided to further strengthen the validity claims for the test and asked HumRRO to conduct a criterion-related validation of the test.

The field test relied on content validity as the basis for establishing a relationship to the VSR position. Content validity asks the question: How well does the assessment sample the

range of important tasks, behaviors, or knowledge associated with effective job performance? Legal and professional authorities (*Uniform Guidelines on Employee Selection Procedures*; Equal Employment Opportunity Commission, 1978 and the *Principles for the Validation and Use of Personnel Selection Procedures*; 4th ed., Society for Industrial and Organizational Psychology - SIOP, 1999) have converged on several basic principles for content validation studies. Evidence for content validity comes from following well-established and accepted job analysis and test development steps and from data that demonstrate a direct link between the selection procedures and job requirements. This is accomplished by: (a) detailing job tasks and the knowledges, skills, and abilities (KSAs) required to perform those tasks; (b) establishing linkages between the job tasks and KSAs and, (c) demonstrating linkages between KSAs and test content areas. In developing assessment instruments, including certification tests, it is HumRRO's practice to follow the guidelines for establishing the content validity of a test even if we plan to use a criterion-related validation strategy, so we had already done the work to establish content validity.

While content-related validation is established through expert judgments, evidence for criterion-related validity consists of demonstrating a useful relationship between a selection procedure (predictor) and one or more measures of job performance (criteria). This is accomplished by administering the predictor tests (i.e., the certification test) to candidates and gathering information on how these individuals perform on the job. Ideally, we would find that individuals who score higher on the tests are those persons who perform more effectively on the job, while individuals who score lower on the tests perform less effectively on the job. The *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 1999) outlines several conditions that should be met before proceeding to conduct a criterion related validity study. They are as follows:

1. Criterion related validity studies should be conducted for jobs that are reasonably stable and are not in a period of rapid evolution.

2. Relevant, reliable, and uncontaminated criterion measures against which to validate the predictor tests are essential for successful criterion related validation studies.

3. The sample on which data are collected should be reasonably representative of the population to which the results are to be generalized.

4. A criterion related validity study should have adequate statistical power to yield a significant predictor-criterion relationship, if one exists. Factors affecting statistical power include sample size, degree of variability in the predictor (i.e., certification test score), reliability of the criterion, etc.

At the time of the validation study, revisions to the VSR position under the CPI model had been in effect at VBA for over a year, and all candidates for certification had been on the job for at least one year. Incumbents had sufficient time to acclimate to the job redesign, so the job was considered stable. We developed a performance measure that combined existing data on productivity and quality with supervisor ratings of performance. This measure met the criterion for relevance described in point 2 and demonstrated sufficient reliability. The sample on which the data were collected included almost 700 candidates, so the sample size was adequate to generalize to the general population of GS-10 VSRs. These factors made criterion-related validity an appropriate strategy for the VSR Certification Test. Results of the criterion-related validity study indicated a strong statistical relationship between scores on the certification test and the measures of job performance.

The May 3, 2006 Test Administration

The first regularly administered test for the Veterans Service Representative (VSR) Certification Program was conducted May 3, 2006. Stations that could not accommodate all candidates in one day also tested on the following two days, as necessary. The test was administered to 934 candidates. Two forms of the test were administered so that different examinees did not necessarily get the same questions in the same order. Each test form included 100 scored questions and 20 additional questions being pilot tested for future use. In the May 2006 test, the two test forms had 67 operational items in common, albeit in different locations within the test. Because of the length of the test, the test is split across two sessions—morning and afternoon. Candidates received a separate booklet for each session.

Due to a processing error at HumRRO, some of the questions in the afternoon booklets were inserted into the incorrect forms. This error resulted in duplicating some questions from the morning session in the afternoon session booklets for the corresponding test form. Quality control procedures in effect at the time included a review of each test booklet, but did not include

a comparison of the morning and afternoon booklets for each test form. Consequently, the processing error was not caught prior to the test administration.

Calls from the field alerted VBA and HumRRO to a potential problem. HumRRO staff investigated to find out how widespread the problem was and alerted VBA to the extent of the problem. Thirty-three items had been duplicated on one form and 34 on the other. The VBA Eastern Area Director, Jim Whitson, set up a teleconference with the HumRRO Project Director and the management members of the VSR Design Team. Subsequently, he sent an announcement to all stations advising the candidates to continue taking the test with the duplicate items assuming that all items would be scored, and that an equitable solution to the problem would be identified as quickly as possible. VBA also made the decision to continue the test as scheduled on the following days, instructing candidates to answer the duplicate items as carefully as though they would be scored. While we had not determined a plan of action, it was possible that we would decide to score the duplicate items, so it was important that candidates answer those items to the best of their ability.

How the Problem Was Handled

On May 11, 2006, HumRRO Vice President Beverly Dugan, VSR Certification Project Director Patricia Keenan, and I met with VBA leadership to discuss the problem and identify possible methods of providing valid scores to participants. Our discussion identified several possibilities, including using some of the pilot items to construct an 80-item test, ignoring the redundancy and scoring each of the duplicate items to provide a 100-item test, and conducting a supplemental administration using the items that were originally intended to be included in each of the afternoon test booklets.

The solution agreed upon was to conduct a supplemental test and administer the items that were originally intended to be presented in each of the afternoon tests. This allowed everyone to be scored on 100 separate items, kept the test mapped to the blueprint exactly as designed, and made the May 2006 administration much more equivalent to the operational field test and the validity test, and to those planned for the future.

The supplemental test was held on June 7, 2006. A total of 46 people who took the May test chose not to sit for certification in the supplemental test; all individuals who chose not to take the supplemental test had failing scores based on the items they did take. The original and

supplemental test questions were scored as intended and 370 (42%) of those who took the entire test passed. The supplemental testing created some inconvenience to the examinees and additional burden to those who administered the tests, but the end result was an assessment that covered the content framework as intended with questions and scores that were psychometrically sound.

Contributing Factors

Several factors contributed to the error in assembling the May 2006 test booklets. One such factor was the limited time available for assembling and checking the test booklets. The VSR job continues to evolve. For example, new types of cases are often added to the caseload, new electronic tools and databases are developed, and more pension cases are being moved to Pension Maintenance Centers. In addition, one of the prime references, M21-1, is undergoing a major revision. HumRRO must rely on expertise of VBA staff members to consider how each new change might affect the validity of the test questions in the VSR certification item bank. A workshop to review test questions was held in April. The item writers reviewed all of the items, revised many of them, and updated the references. Following the workshop, HumRRO staff implemented the edits to the item bank. The revisions were more extensive than anticipated and the work was completed late in the week prior to the scheduled packing date. We had only two days to select the items and put together the four test booklets. The item selection was made more difficult by the fact that, in the two years since the previous administration, many items had become outdated, requiring revision and a new field test, so there were a limited number of remaining items to choose from in some areas of the blueprint. *In retrospect, it was clear that more time was needed for assembling and checking the test forms.*

HumRRO staff members routinely check test booklets for potential problems (e.g., stray marks from the printing process, items split across pages). We did not explicitly compare morning and afternoon versions of the test, which was the only way to have identified the problem. Additional review by VBA experts would be required to provide one additional check of the technical accuracy of each question and the correctness of the scoring key. While scoring was not an issue in the May administration, *it is clear that a more definitive process for final technical review of each test form is needed.*

Preventing the Problem in the Future

First, we have expanded final test form quality control procedures to review the morning and afternoon booklets for each form together. In addition, to relieve the time problems experienced in May 2006, we have changed the timing of the item writing workshops to provide more time after the workshops for assembling and checking operational test forms.

The second problem, the need for a definitive review by VBA experts, will be solved by including reviews of the test items and booklets by Compensation and Pension (C&P) Services staff at VBA. HumRRO will identify the test items to be included in the test and send them to C&P to review the items, keyed responses, and references. After that review, HumRRO will make any needed edits, put together the actual test booklets and send them to C&P for a final review. We implemented this procedure for the August 9, 2006 test and there were no problems with the test.

Summary and Conclusions

The VSR certification test is an important tool for improving the effectiveness of the VSR workforce in serving the benefit needs of our veterans. The testing process is based on a solid analysis of the VSR position and questions were developed and mapped to an established blueprint derived from that analysis. The validity of the test scores for making promotion decisions is supported by both content-related and criterion-related validity evidence.

A number of factors contributed to an error in assembling test booklets for the May 2006 administration of the VSR certification test. Once discovered, corrective action was taken that led to appropriate scores computed from test questions matching the design blueprint exactly. We have no reason to question the validity of these scores. Test assembly and review procedures have been expanded to preclude similar errors with future test forms.

References

Equal Employment Opportunity Commission. (August 25, 1978). Uniform Guidelines on Employee Selection Procedures. *Federal Register*, 44, 38290-38315.

Society for Industrial and Organizational Psychology (SIOP). (1999). *Principles for Validation and Use of Personnel Selection Procedures.* (Fourth Edition). College Park, MD: Author.