Preface

The following is a compilation of all major validation studies involving the Office Proficiency Assessment and Certification System (OPAC). This compilation does not include instructions for operating the OPAC System, and users should refer to either the OPAC Candidate Manual or the OPAC Administrator Manual for such information. The validation studies presented in this compilation date from 1989 to 2005 and certain older validity reports may contain information that is no longer relevant as the OPAC System has been updated and improved over time. Unless otherwise stated, all material presented in this compilation is copyrighted © by Biddle Consulting Group, Inc.

OPAC was distributed by Biddle & Associates, Inc. until 2001. It is currently distributed by Biddle Consulting Group, Inc., which was formed out of Biddle & Associates, Inc. in 2001.

Correspondence should be directed to:

  Biddle Consulting Group, Inc.
  Attn: James E. Kuthy, Senior Consultant
  193 Blue Ravine Road, Suite 270
  Folsom, CA 95630
  Toll-free: 800-999-0438
  Voice (916) 294-4250 · Fax (916) 294-4255
  www.biddle.com

This paper was last updated in January, 2006
# Table of Contents

**Office Proficiency Assessment and Certification®** .......................................................... 9

- Certification Information ........................................................................................................... 10
- Standards for Certification ........................................................................................................... 10
- Procedures for Certification ....................................................................................................... 12
- OPAC Research .......................................................................................................................... 12
- Content Validity Study .............................................................................................................. 13
- Literature Review ...................................................................................................................... 13
- Competency Development ......................................................................................................... 13
- Survey Work ............................................................................................................................. 14
- Data Analysis and Competency Modification ........................................................................... 14
- Results and Reporting ................................................................................................................ 14
- Content Development ............................................................................................................... 15
- Field Test ................................................................................................................................... 15
- Ongoing Research ...................................................................................................................... 15
- OPAC (Entry-Level) Content Validity Study: Areas, Competencies, and Tasks ......................... 15

**Validation Study for Secretarial/Administrative Classifications Using Computer-Based Testing (1991)** .............................................................. 23

- Abstract ..................................................................................................................................... 24
- Method ...................................................................................................................................... 25
- Job Analysis - Part I ..................................................................................................................... 26
- Criterion Development and Sampling ...................................................................................... 26
- Job Analysis - Part II, and Testing of Incumbents on Computers .............................................. 26
- Results ...................................................................................................................................... 27
- Content Validity ......................................................................................................................... 27
- Concurrent Criterion-Related Validity ....................................................................................... 27
- Alternative Procedures Investigated ........................................................................................ 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion</td>
<td>29</td>
</tr>
<tr>
<td><strong>Content and Concurrent Criterion-Related Validity for Some OPAC® Tests</strong></td>
<td>31</td>
</tr>
<tr>
<td>Introduction</td>
<td>31</td>
</tr>
<tr>
<td>Introduction</td>
<td>32</td>
</tr>
<tr>
<td>Experimental Test Battery</td>
<td>32</td>
</tr>
<tr>
<td>Identification of Sample</td>
<td>32</td>
</tr>
<tr>
<td>Job Performance Ratings</td>
<td>33</td>
</tr>
<tr>
<td>Data on the Experimental Tests</td>
<td>33</td>
</tr>
<tr>
<td>Job Analysis and Test Evaluation</td>
<td>33</td>
</tr>
<tr>
<td>Content Validity Results</td>
<td>34</td>
</tr>
<tr>
<td>Concurrent Criterion-Related Validity Results</td>
<td>35</td>
</tr>
<tr>
<td>OPAC Tests</td>
<td>35</td>
</tr>
<tr>
<td>Correlations Between OPAC Test Scales and Supervisory Ratings</td>
<td>36</td>
</tr>
<tr>
<td>Alternative Procedure Analysis</td>
<td>37</td>
</tr>
<tr>
<td>Overall Conclusions Considering Validity and Adverse Impact</td>
<td>38</td>
</tr>
<tr>
<td><strong>Content and Criterion-Related Validity Report for the OPAC® System (1994)</strong></td>
<td>39</td>
</tr>
<tr>
<td>Experimental Test Battery</td>
<td>40</td>
</tr>
<tr>
<td>Identification of Sample</td>
<td>40</td>
</tr>
<tr>
<td>Job Performance Ratings</td>
<td>41</td>
</tr>
<tr>
<td>Data on the Experimental Tests</td>
<td>41</td>
</tr>
<tr>
<td>Job Analysis and Test Evaluation</td>
<td>42</td>
</tr>
<tr>
<td>Content Validity Results</td>
<td>42</td>
</tr>
<tr>
<td>Concurrent Criterion-Related Validity Results</td>
<td>43</td>
</tr>
<tr>
<td>Criterion-Related Validity Correlations</td>
<td>43</td>
</tr>
<tr>
<td>Correlations Between OPAC Test Scales and Supervisory Ratings</td>
<td>44</td>
</tr>
<tr>
<td>Alternative Procedure Analysis</td>
<td>46</td>
</tr>
<tr>
<td>Overall Conclusions Considering Validity and Adverse Impact</td>
<td>47</td>
</tr>
</tbody>
</table>
Development Report for OPAC® 5.3 Legal and Medical Transcription Tests

Background
Early Development
Industry Experts
Test Descriptions
Testing Site
Method
Legal Participants
Medical Participants
Materials
Procedure
Results
Legal Transcription
Medical Transcription
Angoff Scores
Performance Differentiation
Job Duty/KSA Linkage
Discussion
References
Development Report for OPAC® PowerPoint Test (2002) ........................................... 95
Development Report for OPAC® Basic Word Test (2005) ..................................... 107
Development Report for OPAC® Basic Excel Test (2005) ..................................... 112
Development Report for OPAC® Contemporary Keyboarding Test (2005).......... 117
Office Proficiency Assessment and Certification®

Certification Standards

Project and Development Team:

Susie H. VanHuss, Ph.D. Project Director
Richard J. Rovinelli, Ph.D. Project Systems Analyst
Carolyn S. Hayes, B.S., CPS Project Coordinator

L. Joyce Arntson, MBA Instructional Developer
Anne L. Matthews, Ed.D. Instructional Developer
Elizabeth W. Tweeten, Ph.D. Instructional Developer

This section originally Copyrighted © 1989, 1990

By
International Association of Administrative Professionals®
(Formerly Professional Secretaries International®)
10502 NW Ambassador Drive
Kansas City, MO 64195-0404
816 891-6600

All Rights Reserved
Certification Standards

Candidates who take all required modules and units of the OPAC program and meet the standards specified in this section are offered certification by Professional Secretaries international (PSI)--since 1942, the world’s leading organization for office professionals.

Certification Information

Certification provides benefits to both candidates who earn the certification and organizations that support candidates in their bid to earn certification. Candidates gain prestige from certification by the recognized international association for the office profession. Certification enhances personal satisfaction and builds self-confidence. It provides an incentive to continue career development. In addition, candidates receive objective information about their strengths and weaknesses that helps them to formulate realistic plans for career growth. Organizations benefit from the increased professionalism of its entry-level employees. Certification helps to establish a standard barometer for competency within the industry and provides incentive for career growth.

Candidates who have taken all required modules and units of the OPAC program and who have met the standards specified in the next section may apply for certification. The application process consists of having the test administrator export the data from the hard disk of the system to a blank floppy diskette. The diskette must be sent to the OPAC Support Office, 410-C Veterans Road, Columbia, SC 29209 with a check made payable to PSI for $30. Procedures for exporting data are provided in the OPAC System Installation Manual. A form is provided to facilitate the certification process. The OPAC Support Office verifies that the standards have been met and notifies PSI. The certification is then issued by Professional Secretaries International. For additional information about certification, write Professional Secretaries International, P.O. Box 20404, Kansas City, MO 641950404 or call (816) 891-6600, Extension 238.

Standards for Certification

Candidates who wish to receive certification from Professional Secretaries International must meet the standards specified in the next section of this manual. Candidates may repeat those modules and units on which they did not meet certification standards. The OPAC system stores and maintains the response data and results the first time the candidate takes each unit. When units are repeated, the system maintains the response data and the results of both the initial time and the most recent time they have taken the examination. Therefore, once candidates have completed any units successfully, they should repeat only those units in which their scores did not meet the standards specified.

Module 1

The candidate must key at least 45 gross words per minute on the five-minute timed writing with a maximum of 10 errors.
**Module 1, Unit 2**

The candidate must demonstrate the ability to use all of the following word-processing functions:

<table>
<thead>
<tr>
<th>bold</th>
<th>block indent</th>
<th>center</th>
</tr>
</thead>
<tbody>
<tr>
<td>copy</td>
<td>delete</td>
<td>hard hyphen</td>
</tr>
<tr>
<td>hard page break</td>
<td>hard return</td>
<td>hard space</td>
</tr>
<tr>
<td>insert</td>
<td>move</td>
<td>spell check</td>
</tr>
<tr>
<td>printing</td>
<td>underscore</td>
<td>windows/orphans</td>
</tr>
</tbody>
</table>

**Module 1, Unit 3**

The candidate must select the appropriate paragraphs and merge them, and the letter must be formatted correctly in the style indicated. The letter is checked to determine that the proper paragraphs were selected, all appropriate parts of the letter are included, and the positioning of the letter parts is appropriate. The standard is 100 percent. The document is either correct or incorrect.

**Module 1, Units 4, 5, and 6**

The standard for these three combined units is 70 percent. This standard is applied to the last half of Module 1 (Units 4, 5, and 6) rather than on a unit-by-unit basis. A candidate who has scored an average of 70 percent of the three units will be certified.

Unit 7 of Module 1 is not required for certification.

**Module 2, Unit 1**

The standard for Module 2 Unit 1 is 70 percent.

Units 2 and 3 of Module 2 are not required for certification.

**Module 3, Units 1, 2, 3, and 4**

The standard for the entire module is 70 percent. The standard is applied to the total module rather than on a unit-by-unit basis.

**Module 4, Units 1 and 2**

The standard for the composite of these two units is 70 percent. The standard is applied to the combined score rather than on a unit-by-unit basis.

Unit 3 of Module 4 is not required for certification.
Module 5, Units 1, 2, and 3

The standard for the composite of these three units is 70 percent. The standard is applied to the combined score rather than on a unit-by-unit basis.

Units 4, 5, and 6 of Module 5 are not required for certification.

Repetition of Modules

Candidates who do not successfully meet the standards specified on all modules on the assessment may repeat those modules that were not successfully completed. PSI recommends that candidates do additional preparation and/or practice before repeating the modules. The tutorial (OPAC Special Version) should be used for practice in-taking the assessment before repeating the actual assessment.

PSI does not limit the number of times a candidate may repeat the entire assessment or any unit of the assessment. PSI does recommend to test administrators that candidates be allowed to take the assessment three times. Only those modules that were not successfully completed need to be repeated.

Procedures for Certification

Candidates who believe they have met the standards on all units required for certification should have the test administrator extract the test results from the hard disk and export the data to a floppy diskette. The diskette must be sent to the OPAC Technical Support Office for verification that standards have been met. The detailed procedures and a transmittal form for accomplishing this task are contained in the Test Administrator Manual.

The transmittal form and data diskette become the candidate's application for certification. After the scores have been verified, the OPAC Technical Support Office forwards the application to PSI headquarters and notifies PSI that the candidate has met all standards for certification. PSI then issues the certification.

The standard, nonrefundable fee for processing applications, verifying results, and certifying candidates is $30 for each candidate. The certification fee must accompany the application for certification. The check must be made payable to Professional Secretaries International. The candidate's name and identifying number (Social Security Number for U.S. candidates or Canadian National Insurance Number for Canadian candidates) should appear on the check as well as on the transmittal form.

Candidates should not apply for certification until they have met the standards on all required modules and units. The OPAC system captures and maintains data for the initial try and for the most recent repetition of all modules and units.

OPAC Research

Research for the OPAC program is segmented into three phases. The initial phase consisted of a two-year content validity study sponsored by Professional Secretaries International (PSI) that defined the domain of knowledge, skills, and abilities of entry-level office employees and provided information concerning the positions of entry-level office employees. The second phase consisted of developing and field testing the instruments.
used to assess the competencies identified in the content validity study. The final phase is an ongoing research component that will analyze all data collected during a three-to-five year period of use of the assessment in practical settings.

**Content Validity Study**

The validity study was organized into five components:

1. Literature review
2. Competency development
3. Survey work
4. Data analysis and competency modification
5. Reporting

A brief review of each phase follows.

**Literature Review**

The purpose of the literature review was to provide a starting point for the competency development component of the study. A comprehensive literature search provided numerous articles and research studies written in the past five years dealing with competencies needed by secretarial employees, word processing employees, and employees in general office/clerical-type positions.

The major studies which identified and validated a list of specific competencies needed by entry-level workers included DACUM (Developing A Curriculum) studies; V-TECS (Vocational-Technical Education Consortium of States) catalogs of tasks, performance objectives, and performance guides; and studies conducted or sponsored by state departments of education. The remainder of the studies consisted primarily of masters theses, doctoral dissertations, and studies by independent researchers.

The literature review produced a massive list of competencies that had been identified as essential for office employees. This list provided the starting point for the competency development component of the PSI study.

**Competency Development**

The first phase of the competency development process consisted of hiring a content expert to develop an initial set of competencies, knowledges, skills, and abilities utilizing the list of competencies obtained in the literature review. Duplicate competencies were eliminated and similar competencies were combined.

The second phase of the competency development process was an iterative process of writing, reviewing, and refining the competencies. Managers, business educators, and secretaries who were members of the Institute for Certifying Secretaries participated in this phase. A psychometrician was employed to facilitate the group discussion. This synergistic process was used to help validate relevancy of each competency, ensure that the scope of the domain of entry-level knowledges, skills, and abilities was adequately covered; ensure that the competencies were clearly and accurately presented; and organize the competencies into meaningful content dimensions.
The third phase of the competency development process was the review by the Institute Task Force on Entry-Level Certification. This second group of managers, business educators, and secretaries reviewed and refined the competencies. The Task Force was also given oversight responsibility for the study. The resulting product of the competency development component was a list of 49 competencies that were organized into eleven job-content dimensions. These competencies were then used to develop surveys that were administered to random samples of secretaries, business educators, and managers.

**Survey Work**

This component consisted of two surveys and a job function diary. Each survey was mailed to a random sample of members of Professional Secretaries International, business educators, and managers. Participants provided ratings on the importance of each of the 49 competencies as well as the frequency in which the competency would be used by an entry-level person, and whether or not the competency represented an essential skill, knowledge, or ability. Participants also provided both importance ratings and an estimate of the percentage of time an entry-level person would spend in each of the eleven job dimensions. Bio-demographic data were also collected.

To obtain more data to augment the "essential/non-essential" data for the study, a job function diary study was conducted. The purposes of the diary study were to:

1. To determine what tasks and skills are performed by entry-level personnel during specified work periods.
2. To determine if size of organization makes a difference in the types of tasks required of entry-level personnel.
3. To determine if the tasks and skills identified were covered by an existing entry-level competency.

**Data Analysis and Competency Modification**

A cyclical process was used to integrate this component with the survey work. Data were analyzed and reviewed after each survey and after the job function diary study. Competencies were modified based on the survey work.

**Results and Reporting**

The final report of the content Validity study was prepared and presented to the Institute Task Force on Entry-Level Certification for approval. The report consisted of the survey results and an approved, validated set of competencies that were later used in the development of the entry-level examination program.

The list of competencies was comprehensive as judged by the results of two surveys, as well as by the efforts of members of the Institute for Certifying Secretaries and the Institute Task Force on Entry-Level Certification. Of the 49 competencies, 36 were considered essential for the successful performance of an entry-level office employee. In addition to the delineation of the essential competencies, the study also provided specific information on the importance and frequency of use of these competencies.
Content Development

The content validity study defined the content domain in terms of the knowledge, skills, and abilities required for successful performance as an entry-level office employee. A total of 36 competencies were identified as essential for successful performance.

The relevancy (importance) of each competency and the representativeness (frequency) of each competency were identified. These data served as the basis for the examination blueprint and specifications. For a list of the specific competencies, refer to the section of the manual entitled, Competencies.

Field Test

Over 300 individuals representing over 30 educational institutions and organizations participated in the field test. Field test data were used to make minor content modifications, determine appropriate time frames for the various units, and to set the standards for certification. Information from the technical report is available from Professional Secretaries International.

Ongoing Research

Performance data from candidates who take the assessment will be collected and analyzed over a three-to-five year period. The ongoing research component will be used to determine the extent to which the assessment data meets employment testing standards and to study the relationship between results on the assessment and job performance.

OPAC (Entry-Level) Content Validity Study: Areas, Competencies, and Tasks

Status Codes:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Assessed in the Office Proficiency Assessment and Certification Program.</td>
</tr>
<tr>
<td>N</td>
<td>Competencies that were identified in the Content Validity Study as not being essential for entry-level employees.</td>
</tr>
<tr>
<td>E</td>
<td>Competencies that were identified as essential, but that are not assessed in the OPAC program at this time. Exploratory work is being done for future assessment.</td>
</tr>
</tbody>
</table>

Company Organization and Policies

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1.0 Is knowledgeable about the products or services of the company.</td>
</tr>
<tr>
<td>N</td>
<td>1.1 Is knowledgeable about the organizational structure of the company.</td>
</tr>
<tr>
<td>N</td>
<td>1.2 Is knowledgeable about company policies, both formal and informal.</td>
</tr>
</tbody>
</table>
### Technology in the Office

<table>
<thead>
<tr>
<th>Code</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.0</td>
<td>Is knowledgeable about technological changes and innovations and their impact on a business office.</td>
</tr>
<tr>
<td>A</td>
<td>2.1</td>
<td>Understands data and information processing concepts and is familiar with the basic terminology relating to data and information processing.</td>
</tr>
<tr>
<td>A</td>
<td>2.2</td>
<td>Understands the basic concepts of telecommunications such as electronic mail, facsimile communications, etc., and their impact on distributing information.</td>
</tr>
<tr>
<td>A</td>
<td>2.3</td>
<td>Understands the role information processing plays in an office information system and knows terminology common to information processing.</td>
</tr>
<tr>
<td>A</td>
<td>2.4</td>
<td>Understands the role of computers in an office information system and is able to utilize the word-processing function of computers.</td>
</tr>
<tr>
<td>N</td>
<td>2.5</td>
<td>Is familiar with different types of information equipment and systems and understands how various software packages can extend the capacity of the information processing equipment.</td>
</tr>
<tr>
<td>A</td>
<td>2.6</td>
<td>Is able to follow general instructions for operating information-processing equipment.</td>
</tr>
<tr>
<td>A</td>
<td>2.7</td>
<td>Understands the function of and is able to operate printers.</td>
</tr>
</tbody>
</table>

### Human (Interpersonal) Relations

<table>
<thead>
<tr>
<th>Code</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>3.0</td>
<td>Realizes the importance of developing and promoting good human relations and is aware of His/her role in relation to superiors, peers, subordinates, clients or customers, and sales or service personnel.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Displays an understanding and acceptance of himself/herself.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Recognizes the needs and personal characteristics of others with whom he/she works.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Recognizes the importance of working cooperatively and getting along with others.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. Demonstrates tact in sensitive and/or difficult situations.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>e. Conducts his/her office activities in an ethical manner.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>f. Exhibits consideration and respect for others in the workplace.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>g. Develops and maintains a positive work attitude and exhibits responsible work behavior.</td>
</tr>
<tr>
<td>E</td>
<td>3.1</td>
<td>Recognizes that effective career planning and career advancement require that the objectives of the individuals must be compatible with the objectives of the organization.</td>
</tr>
<tr>
<td>E</td>
<td>3.2</td>
<td>Is able to communicate clearly with employers, fellow workers, and people outside the company both orally and in writing.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Understands the communication process and its value in human and business relations.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Recognizes some of the problems in maintaining effective communications.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Strives to improve communications by improving listening skills, using direct simple language, utilizing feedback, and timing messages carefully.</td>
</tr>
</tbody>
</table>
### Basic Office Skills

**A 4.0** Demonstrates efficient and effective ways of organizing his/her time to complete work assignments.

- a. Analyzes the jobs he/she performs daily and devises a structured plan in order to reduce or eliminate wasted time and increase productivity.
- b. Uses good judgment and careful thinking; establishes priorities for handling the work assigned.
- c. Demonstrates the ability to use timesaving procedures and devices.

**A 4.1** Maintains a well-organized work station to insure a smooth flow of work.

**A 4.2** Performs document-producing tasks and keyboarding functions using a variety of information processing equipment.

**A 4.3** Operates information processing equipment to record, edit, print, store, and revise correspondence, reports, statistical data, forms, lists, and other materials. This equipment includes automatic typewriters, text-editing machines, transcription machines, printers, OCRs, and other equipment that extends information processing capabilities.

- a. Exhibits expert keyboarding skills essential to document-producing tasks.
- b. Understands and is able to use all special features of information processing equipment such as merge, pagination, etc.

**A 4.4** Produces mailable business communications and carries out instructions from manual or machine dictation.

- a. Keyboards from both longhand and typewritten rough drafts, pre-recorded dictation, and machine dictation.
- b. Edits rough drafts and unarranged copy for proper punctuation, paragraphing, grammar, etc.
- c. Knows basic operating procedures for transcription machines and uses proper machine transcription techniques.
- d. Uses listening and decision-making skills when transcribing from machine transcription.
- e. Is able to follow special instructions for dictated materials and uses effective techniques of planning, transcribing, and distributing the work.
- f. Selects proper stationery and plans the proper format for assigned tasks.

**A 4.5** Utilizes basic business knowledge, skills, and vocabulary in processing work.
<table>
<thead>
<tr>
<th>Assignment</th>
<th>Description</th>
</tr>
</thead>
</table>
| A 4.6      | Is able to accept responsibility and to carry out assignments with limited guidance or supervision.  
  a. Grasps and follows instructions quickly.  
  b. Organizes materials and supplies for efficient handling and uses equipment and resources effectively.  
  c. Meets expected deadlines within a regular working day (except in unusual situations). |
| N 4.7      | Knows company standards and procedures for processing documents.  
  a. Follows company procedure manuals.  
  b. Uses standard company formats and is able to adapt standard formats to special situations.  
  c. Meets established quality standards and production deadlines. |
| A 4.8      | Exhibits a high level of mental concentration and demonstrates the ability to work under pressure of production requirements. |
| E 4.9      | Is able to select and purchase appropriate stationery; typewriter, filing, and mail supplies; desk accessories; and other office supplies.  
  a. Identifies and keeps a file on all sources of office supplies.  
  b. Prepares all requisitions, purchase orders, and/or invoices for replenishing office supplies.  
  c. Develops a procedure for maintaining the proper inventory level of all supplies.  
  d. Maintains an orderly supply cabinet with supplies arranged conveniently for general use. |
| A 4.10     | Is responsible for the reproduction of all types of typewritten and printed documents.  
  Is familiar with the different reprographic processes and is able to select the appropriate process for the given situation.  
  Prepares materials to be photocopied and is able to operate his/her firm’s copying machines.  
  Prepares requisitions and instructions for materials to be reproduced.  
  Knows copyright guidelines and follows them in making decisions about legal or illegal copying. |
| N 4.11     | Assists the executive and other professionals in gathering, processing, and verifying information needed for preparing reports, presentations, manuals, and other publications.  
  a. Knows what reference sources are available and how to use those resources.  
  b. Gathers data from resource documents and research materials and organizes data into a usable form. |

**Language Arts Skills**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 5.0</td>
<td>Applies basic language arts skills in the composition and keyboarding of all documents.</td>
</tr>
</tbody>
</table>
a. Knows correct grammar.
b. Knows how to spell words commonly used in business.
c. Knows how to punctuate correctly.
d. Knows how to use capitalization effectively.
e. Knows how to use possessives properly.
f. Knows rules-for correct number usage.
g. Knows how to use abbreviations correctly.

A 5.1 Carefully checks all documents for accuracy.

a. Proofreads letters, memos, and reports for correct grammar; punctuation; spelling; logical, clear content; and correct and complete data.
b. Proofreads statistical copy for accuracy and adds columns of figures if a total is given.

Mail, Telephone, and Appointments

A 6.0 Processes mail quickly and efficiently.

a. Sorts, opens, date stamps, prioritizes, and distributes mail to specified individuals and/or departments.
b. Expedites the executive's handling of the mail by providing background information and/or pertinent files where appropriate.
c. Keeps a mail register when required by company policy.
d. Prepares outgoing mail so that it can be processed quickly and accurately by the Postal Service.
e. Addresses envelopes in accordance with Postal Service rules.
f. Includes all enclosures and folds and inserts letters properly in the envelopes.
g. Knows the different classes of mail and the special mail services available so that he/she can determine the appropriate class to be used on outgoing mail.
h. Is familiar with and practices ways to reduce mailing costs.
i. Is able to handle any special problems that arise in processing mail, i.e., mailing currency, retrieving mail incorrectly addressed, changing addresses, forwarding mail, etc.
j. Is familiar with international mail regulations and services.
k. Is knowledgeable about other mailing and shipping services and is able to make decisions about other means of shipment based on cost, speed of delivery, and convenience to shipper and receiver.

A 6.1 Has knowledge of telephone services and is able to handle telephone duties skillfully.

a. Takes appropriate action in given situations, i.e., handling problem calls, putting callers on hold, transferring calls, placing long distance calls, etc.
b. Uses appropriate techniques in placing and receiving telephone calls promptly and efficiently.
c. Develops a good telephone personality and uses proper telephone etiquette.
d. Records telephone messages completely and correctly and delivers them promptly.

A 6.2 Is responsible for scheduling appointments, maintaining office calendars, and receiving office callers.
a. Maintains a business-like office atmosphere and exhibits professional behavior when receiving callers and scheduling appointments.
b. Follows the employer's preferences when scheduling appointments.
c. Records complete information regarding date, time, place, purpose, and participant's when scheduling appointments.
d. Coordinates his/her appointment calendar with that of his/her employer.
e. Keeps a record of office callers and refers them to the proper person(s).

### Written Communications

<table>
<thead>
<tr>
<th>A</th>
<th>7.0</th>
<th>Composes routine business documents (letters, memos, reports, etc.) and presents them in a clean, error-free typewritten format that is consistent with accepted business practices and styles.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Knows the characteristics of an effective business letter and includes those characteristics in composing business letters.</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Plans the letter before composing, i.e., gathers all the facts, determines what must be included in the letter, decides upon the order of presentation, and develops an effective beginning and ending.</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Identifies the different parts of a letter and knows their correct placement within the letter.</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Selects appropriate salutations and complimentary closes.</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>Formats documents appropriately.</td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>Has knowledge of different types of business reports and is able to prepare them according to accepted styles and formats.</td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>Differentiates between formal and informal reports.</td>
<td></td>
</tr>
<tr>
<td>h.</td>
<td>Knows the parts of different kinds of reports and is able to arrange them properly within the report.</td>
<td></td>
</tr>
<tr>
<td>i.</td>
<td>Knows how to construct and format charts and tables.</td>
<td></td>
</tr>
<tr>
<td>j.</td>
<td>Is familiar with the commonly used business forms in his/her office and is able to locate the information necessary to complete the forms and to fill in that information correctly.</td>
<td></td>
</tr>
</tbody>
</table>

### Records Management

<table>
<thead>
<tr>
<th>A</th>
<th>8.0</th>
<th>Understands the principles of records management.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8.1</td>
<td>Is able to arrange business records in accordance with a systematic plan and file them in such a manner that they can be located quickly.</td>
</tr>
<tr>
<td>a.</td>
<td>Identifies basic filing methods and determines the best filing method for the active records.</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Knows and applies basic filing rules.</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Prepares records for filing by indexing, coding, sorting, and (if necessary) cross-referencing.</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Prepares folders and labels for records to be filed.</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>Maintains the confidentiality of records under his/her responsibility.</td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>Determines the effectiveness of existing filing systems and makes recommendations for reorganization where applicable.</td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>Is familiar with filing supplies and equipment and makes recommendations or provides for the acquisition of such.</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>8.2</td>
<td>Assists users in the retrieval and use of records.</td>
</tr>
</tbody>
</table>
### Financial Records

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>9.0</td>
<td>Retrieves information from financial reports.</td>
</tr>
<tr>
<td>A</td>
<td>9.1</td>
<td>Performs simple accounting tasks and keeps some permanent records.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Establishes and maintains a petty cash fund.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Keyboards financial statements such as balance sheets and income statements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Maintains basic financial records.</td>
</tr>
<tr>
<td>A</td>
<td>9.2</td>
<td>Performs banking activities.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Makes deposits, writes and records checks, endorses checks' and reconciles bank statements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Understands electronic fund transfer, direct deposit or payment telephone transfer, etc.</td>
</tr>
<tr>
<td>A</td>
<td>9.3</td>
<td>Operates office machines that are widely used in computing and producing financial records.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Performs basic math functions on an electronic calculator or other similar machines.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Utilizes office machines to solve business math problems encountered in financial-related tasks.</td>
</tr>
</tbody>
</table>

### Travel

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10.0</td>
<td>Makes business travel arrangements according to company policies and procedures.</td>
</tr>
</tbody>
</table>
a. Knows executive preferences with regard to transportation and accommodations.
b. Sets up a trip file to accumulate all the information and materials relating to a particular trip.
c. Makes hotel/motel and transportation reservations, processes requests for use of company auto or plane, and makes arrangements for travel funds.
d. Prepares a complete itinerary.
e. Notifies associates of the executive's plans to be away from the office and attends to the cancellation and/or rescheduling of meetings scheduled during the executive's absence.
f. Collects and organizes materials that are necessary for the successful completion of the trip.

N 10.1 Assumes responsibility for routine office activities during the executive's absence.

a. Handles daily communications and activities within the scope of his/her authority and refers exceptions to appropriate personnel.
b. Forwards mail if necessary and maintains a file or files of mail and other communications and information that will be held for action until the executive returns.
c. Follows proper procedures for handling the executive's mail during his/her absence.

N 10.2 Is responsible for compiling and preparing expense reports.

a. Collects the receipts and information necessary for completing an expense report.
b. Verifies amounts, dates, and places, and enters the information in the proper form.

Meetings

N 11.0 Assists in the planning, organizing, and implementing of business meetings.

a. Assists in site selection and reserving meeting rooms.
b. Notifies participants of date, time, place, and purpose of the meeting.
c. Prepares and distributes the meeting agenda.
d. Reserves the equipment needed to conduct the meeting and prepares and assembles materials to be used during the meeting.
e. Performs any follow-up activities required after the meeting.

A 11.1 Records, transcribes, and distributes the minutes of the meeting.
Validation Study for

Secretarial/Administrative Classifications


Janet M. Burns
Los Alamos National Laboratory

Los Alamos National Laboratory is a United States Department of Energy (DOE) national laboratory, managed by the University of California.

Abstract

This paper presents the results of a content and concurrent criterion-related validity study conducted at Los Alamos National Laboratory for clerical, secretarial and administrative classifications using computer-based testing. The advantages and disadvantages of different types of testing software incorporated in the study are explored. Job analysis methodology, procedure for establishing cut-off scores and comparative adverse impact and validity are analyzed.
The purpose of this paper is to present the results of a content and concurrent criterion-related validation study conducted on the administrative and secretarial classifications at Los Alamos National Laboratory. The Laboratory retained Biddle & Associates/Biddle Consulting Group Inc. to assist in the design, methodology, form development, analysis and preparation of the initial compliance report. The objective of the study is to replace the Lab's traditional typing test administered on IBM Selectric typewriters with a computer-scored testing environment, including a word processing assessment, for the selection of secretarial and administrative classifications. Phase II of the study will expand the office skills assessed to include spreadsheets and data entry.

Los Alamos National Laboratory has administered a traditional typing test to applicants for secretarial, clerical and administrative positions for over two decades. Test scores are only one of the criteria considered in the selection process. As in many organizations, the technology being applied on the job today is far more advanced and thus has outdated the traditional types of tests used to select secretarial personnel. As this gap continues to diverge, the current test fails to provide the necessary information required to make sound selection decisions. Applicants and hiring managers have requested more state-of-the-art procedures. Our goal is to select and implement a system, which more accurately assesses a broader range of skills including word processing, and provides more in depth information about those skills than just speed and accuracy on a typewriter.

The Lab's administrative population is close to 1200 individuals across 24 job titles. The Lab's total population is approximately 7400. Between 70 and 130 candidates test each month with a significant number of selections made annually for these titles. Historically, applicants have had to pass the typing test at one of two different speeds depending on the requirements of the position in order to progress to the next phase of the selection process. The cutoff score is 55 words per minute and 5 errors for the secretarial test, and 25 words per minute and 5 errors for the clerical test. Early investigation of these job titles and job content indicated that the required office skills varied within and between classifications. This situation becomes even more complex when incumbents use multiple software packages and an organizational standard for word processing software does not exist. These and other factors to be discussed strongly influenced the design of this study. The following sections will explain how we dealt with the uniqueness of this study and the results which followed.

**Method**

This study was conducted following Section 15C of the Uniform Guidelines on Employee Selection Procedures (1978). The primary methodology is content validity. Criterion-related validity was included to augment the study and answer some additional questions. Multiple classifications, multiple kinds of word processing software, variety of required office skills within and between classifications, and the need for a computer-based testing and scoring system greatly influenced our approach.

**Identification of Tests and Test Publishers**

Four test publishers were identified for inclusion in the study based on the types of computer scored tests available, the range of skills which could be assessed, existing validation research, cost, and whether specialized word processing software tests (i.e. WordPerfect, MultiMate, etc.) were available. We were also interested in "generic" clerical tests for individuals without word processing experience and for positions which might
require assessing office skills other than word processing such as editing, grammar or spelling.

MANPOWER INC.\(^1\), Tap Dance\(^2\), QWIZ Inc.\(^3\), Office Proficiency and Assessment Certification\(^4\) (hereinafter called OPAC\(^5\)) submitted tests for the study. A total of 12 tests were selected for this phase and are listed in Appendix A according to the type of test along with a short descriptive footnote of each test. Appendix B lists a total of 41 test scales being measured by the 12 tests. The number of tests, test scales and different word processing software packages complicated the analysis significantly.

**Job Analysis - Part I**

A survey was sent to incumbents of each of the 24 clerical, secretarial and administrative classifications for which the current typing test was being used to see if a word processor was being used, and if so, on what equipment and which software package. 749 of the 1139 incumbents responded to this first survey. WordPerfect, Microsoft Word for the IBM, Microsoft Word for the Macintosh and MultiMate were identified as the software packages used most frequently. The number of users for WordPerfect and Microsoft Word for the Macintosh were 249 and 259 respectively. Only 43 of the 1139 incumbents indicated they were not using word processing on the job and thus were not included in the study. The fact that 96% of the Lab’s population is using word processing confirmed the need for a replacement to the current test which is not measuring this skill.

**Criterion Development and Sampling**

Supervisors of the 749 job incumbents who responded to the first survey were invited to a supervisory workshop. The purpose of the workshop was to identify the skills and levels of skills used in the effected classifications. The form used to gather the data on skill ratings is in Appendix C. The rating scales were developed based on what can be measured by the selected tests. Using a scale from 1-5 supervisor’s were asked to rate the employee’s speed, accuracy, and (where requested) levels of skill for the nine office skills listed. A description of each skill and definitions for each rating scale were provided. Supervisors were instructed to provide ratings only for the skills for which they had first-hand knowledge. As mentioned earlier, the rating form includes skills such as spreadsheet, database management and data entry, which will be part of Phase II.

Over 100 supervisors participated in the criterion workshops resulting in ratings for 292 unique individuals. 259 incumbents received single ratings while 33 received more than one rating. The multiple ratings were averaged for the analysis.

**Job Analysis - Part II, and Testing of Incumbents on Computers**

Incumbents who received ratings by their supervisors were invited to take the tests included in the experimental test battery. Participation was voluntary. At the time the incumbents took the tests, they were asked to complete a form that gathered job analysis and test evaluation information. Incumbents were asked several questions as subject matter experts: is some level of the skill being measured needed, identify a duty for which

---

\(^1\) MANPOWER Inc., is a registered trademark of MANPOWER Inc.
\(^2\) Tap Dance is a trademark of International Testing Services Inc.
\(^3\) QWIZ Inc., is a registered trademark of QWIZ Inc.
\(^4\) OPAC is a registered trademark of Professional Secretaries International.
the skill is required, is the skill distinguishing, is the test a representative sample of the skill used on the job, did the test require more skill from the test taker than is required on the job, can the skill be learned in less than 8 hours, does the test resemble one or more job duties, and does the product of the test resemble a work product. After taking the tests subject matter experts were then asked to estimate what a minimally qualified applicant's score should be on each test scale. Using a 1-5 importance rating scale incumbents were also asked to rate the job duty. A questionnaire was completed for every test administered. This information was captured in several databases. Every question and each test scale were analyzed for the content validity report.

While 113 incumbents participated in the testing, they were only asked to take tests that they linked to one or more job duties. Others did not complete the entire test battery for various reasons. Therefore, the sample sizes for each test are not equal.

Tests were administered on five IBM PC's and one Macintosh SE over a seven-week period. The length of time to complete all 12 tests ranged from 5 to 9 hours depending on the individual.

A survey was also conducted of nine local high schools and community colleges to verify that students were being taught with word processors rather than typewriters. Every school surveyed is using word processors except for the one private school in the sample. WordPerfect was the most prevalent software being used.

Results

Content Validity

Evidence demonstrating that each test scale is a representative sample of a duty performed on the job was established through the content validity questionnaire. A minimum of 50% of the incumbents was set as a minimum standard of acceptance. This means that 50% of the incumbents had to agree on each of the content validity questions described in the methodology. A standard of 70% was set as the preferred standard. Each test scale easily passed all the minimum standards except for Manpower's multiple choice word processing test for WordPerfect users. Incumbent responses indicated that this test requires more skill than is needed on their particular jobs.

Concurrent Criterion-Related Validity

Hypotheses and the anticipated direction were determined for each test scale to each relevant rating scale. A one-tail .05 level of statistical significance was applied with the specified direction. Pearson Product-Moment Correlations were calculated for each of the hypothesized relationships. Two of the test publishers had test scales that were correlated significantly at the 5% level of chance with speed ratings from the supervisors: OPAC and Tap Dance. Both of their 5-Minute Tests demonstrated statistical significance with speed performance ratings and Tap Dance's Word processing Test also showed statistical significance to the speed ratings.

Ratings of accuracy of work performed were correlated significantly with scales from three of the test publishers: Manpower's Ultraskill, OPAC's Language Arts, and Tap Dance's Editing and Word processing tests.
Level 1 word processing skill as evaluated by supervisors was correlated significantly with tests from all four test publishers. Level 2 ratings were correlated significantly with tests from three test publishers: Manpower, OPAC and QWIZ. When the data is analyzed separately by software, sample sizes decrease and statistical significance is not achieved for all software specific correlations. The 5-minute typing tests from OPAC, Tap Dance and the Lab all correlated significantly with speed, while none of the 5-minute tests correlated with accuracy. QWIZ did not predict speed or accuracy though the sample sizes were smaller. Manpower does not measure speed with a 5 minute timed typing test.

**Alternative Procedures Investigated**

Each of the test scales showing statistical significance with supervisory ratings was analyzed through the Biddle Consulting Group Statistical Cutoff Program. The program calculates statistical significance between groups for each score and calculates practical significance as well. According to Section 4D of the Uniform Guidelines on Employee Selection Procedures (1978) both statistical and practical significance must be shown in order for adverse impact to exist.

Cochran's correction to the chi-square was used for statistical significance at the .05 level of probability. Practical significance exists if it took more than adding two people to the disadvantaged group to change the statistical significance finding, more than 3 people added to the disadvantaged group to change the 80% Rule of Thumb test, and more than four people to bring the passing rates of the two groups to within 2.1% of each other. Using these rules, adverse impact was found for one or more scores in the samples for Manpower's spelling scale, and OPAC's words per minute and keystrokes scale. For each of these test scales there were several scores without adverse impact. However, because the incumbents taking these tests had already passed the Los Alamos Typing Test, these results need to be reanalyzed with unrestricted data from applicants in general.

When evaluating test publishers on the basis of statistical validity, content validity and adverse impact against the accuracy rating scale, OPAC and Tap Dance tests produce validity without adverse impact - The Manpower Ultraskill spelling scale does produce adverse impact with validity, while the other scales produce validity without adverse impact. No QWIZ test scale produced statistical validity with the accuracy ratings. When evaluating test publishers on the basis of statistical validity, content validity and adverse impact against the speed rating scale, the OPAC 5-Minute Test produced adverse impact and validity. The Tap Dance 5-Minute Test and Word processing tests produced validity without adverse impact.

When evaluating test publishers on the basis of statistical validity, content validity and adverse impact against the word processing level 1 rating, all test publishers produced validity without adverse impact.

When evaluating test publishers on the basis of statistical validity, content validity and adverse impact against the word processing level 2 rating, all four test publishers produced validity without adverse impact. However, the statistical validity for Tap Dance's Word processing Test was found for only the Microsoft sample.

Tests to test correlations were also performed for samples greater than or equal to ten. A procedure similar to identifying the test to ratings relationships was applied. Each test had many scales. Only those scales that we hypothesized to have the most obvious...
relationships were selected for the analysis. It is possible that correlations could exist between scales that were not analyzed.

**Discussion**

An overwhelming amount of data has been collected and analyzed in this study. This discussion will focus on three important areas that emerged from the analysis: content validity design, equal validities and adverse impact, and the intercorrelations.

The content validity approach used in this study allowed us to validate a number of different tests across numerous secretarial and administrative job classifications where incumbents within a classification are using word processing at different levels. This was a non-traditional approach to job analysis that addresses Section 14C of the Uniform Guidelines on Employee Selection Procedures (1978). The responses to the questionnaires showed outstanding support for almost all of the tests with cutoff scores established at the point which 70% of the incumbents agreed on that score or a more stringent score. Regardless of the test the Lab selects, as job openings occur, hiring managers will have to identify whether or not word processing is a requirement for that position within a classification. By focusing on the common skills the testing function will be more responsive to changing and varied job requirements at the Lab. Job content will drive the process rather than strictly job title.

This study presented a unique illustration of Section 3B, Suitable Alternatives, of the Uniform Guidelines on Employee Selection Procedures (1978). When alternative selection procedures (i.e. different tests or test scales) or alternate uses of a selection test (i.e. different weights within a job-related range or alternate cutoff’s) are substantially equally valid for a given purpose, the one with less adverse impact should be used. The OPAC and Tap Dance 5 Minute Timed Typing Tests each produced statistically significant validities with speed ratings, \( r = .28, n = 62 \) and \( r = .31, n = 57 \), with and without adverse impact, respectively. The correlation between the two tests was \( r = .88, n = 55 \). The validity coefficients are not significantly different. The Tap Dance Word Processing test also predicted speed, \( r = .43, n = 49 \), without adverse impact. The OPAC 5-Minute Timed Typing Test and Tap Dance Word Processing Test validity coefficients are not significantly different. The intercorrelation of \( .47, n = 45 \) is significant.

The Manpower Ultraskill spelling scale and Tap Dance Editing error scale evaluated against accuracy also produced substantially equal validities, \( r = .34, n = 72 \) and \( r = .49, n = 53 \), with and without adverse impact, respectively. The correlation between the two tests was significant, \( r = .36, n = 47 \).

When the Manpower Ultraskill spelling scale and OPAC Language Arts spelling scale are evaluated against the accuracy rating scale, both tests produced substantially equal validities, \( r = .34, n = 72 \) and \( r = .44, n = 39 \), respectively. Only the Manpower spelling test produced adverse impact. Though both are spelling test scales the intercorrelation of \( .19 \) was not significant. It appears each test is measuring different parts of the accuracy criterion.

Correlations between the tests are interesting, however the sample sizes restrict any definitive conclusions. Analyzing the data separately for each specific software unavoidably reduced the sample sizes. Although a major effort was made to have every incumbent take every test this was not always possible. Of the correlations analyzed OPAC has a moderate relationship with Tap Dance and the Manpower RAP written test, and none with the Manpower Ultraskill test. Tap Dance appears to be measuring some of the same skills as
OPAC and the Manpower tests, though the relationship with the written test is stronger. It is the written knowledge test of Manpower that seems to be similar to the other word processing tests. Since the Manpower Ultraskill test is not a "pure" word processing test, and is considered an assessment of clerical skills on a word processor, it is not surprising that minimal or no relationships exist with the other tests. It is not intended to measure "word processing", but is administered on the specific word processing package with which a person must be familiar. There is some relationship however between the two Manpower tests. QWIZ has a very low relationship to OPAC and no relationship with either Manpower test.

The Los Alamos, Tap Dance and OPAC 5-Minute Typing tests predicted ratings of speed but not accuracy. Only Tap Dance measured speed without adverse impact. As hypothesized the 5-Minute Typing Tests show strong intercorrelations with larger samples. All three test publishers and the Los Alamos typing test appear to be measuring a very similar skill. The Los Alamos typing test was very highly correlated to each of the computer-based 5-minute typing tests. Direct restriction of range is present with the Lab's current typing test as well as indirect restriction of range to the extent the others are correlated. When correcting for restriction of range the validity coefficient increases from $r = .3047$ to $r = .395$.

This data is only applicable to the samples used at Los Alamos National Laboratory. All tests included in the study offer their own unique advantages that must be considered along with the statistical results and other practical concerns for each organization. Manpower is the only test publisher with a test for the Macintosh. This is an important issue for the Lab since the number of Mac users is increasing daily. Several criteria will be applied to each of the tests before a decision is made. A predictive study is planned as a follow-up.

**Note:** No reference to this study should imply an endorsement or criticism of the test publishers or their tests.

The author gratefully acknowledges Charlotte Garcia, the Laboratory's Test Administrator for her outstanding work and contribution to this project.
Content and Concurrent Criterion-Related Validity for Some OPAC® Tests

Richard E. Biddle
Introduction

A study was conducted at an employer with more than 5000 employees to examine the validity of several OPAC (Office Proficiency Assessment and Certification) tests. The OPAC tests were originally developed and content validated by Professional Secretaries International. The OPAC tests are computer administered and computer scored.

The employer involved in the study was searching for a word-processing test that could be administered and scored in an independent test environment as was feasible to replace its traditional 5 minute timed typing test. The traditional typing test measured speed and accuracy using IBM selectric typewriters and required extensive test administrator supervision. The employer wanted a test that could measure an applicant's skill at creating, formatting, proofing, and editing documents, while also measuring word-processing skill using a word-processing type program. An applicant's speed and accuracy were also important factors for the test to measure. Also, the employer wanted to minimize the test administrator's time.

More than 20 classifications needed a selection testing procedure that measured keyboarding speed and accuracy as well as some level of word-processing skill.

To add to the problem, many different types of word-processing software were being used. The new selection procedure needed to test applicants using different word processing programs.

Experimental Test Battery

OPAC tests of Language Arts 1, Editing/Formatting from Rough Draft, and Keyboarding were used as part of an experimental test battery.

The OPAC Language Arts 1 test evaluated in the study was used to measure skills in proofing a document to identify errors in grammar, spelling, punctuation, capitalization, possessiveness, number usage, and abbreviations.

The OPAC Editing/Formatting from Rough Draft test was used to measure skills in operating features and functions of a specific word-processing program.

The OPAC Keyboarding test was used to measure an individual's speed and accuracy of typing text on a keyboard.

Identification of Sample

Incumbents of 24 secretarial, clerical, and administrative classifications were sent a survey. The survey asked about the use of word-processing equipment and software. Of the 1139 incumbents who were sent surveys, 65.8% responded (749). The responses showed that WordPerfect, Microsoft Word for the IBM and Macintosh, and MultiMate were the word-processing software most frequently used. Of the 749 incumbents who responded, 94.3% (706) indicated that they were using some form of word-processing on the job. About half (378) used more than one wordprocessor, including text editors or desktop publishing. The 5.7% (43) who used no word-processing on the job were not included further with the study.
Job Performance Ratings

A series of workshops were conducted for those who supervised the survey respondents to obtain ratings of job performance. Supervisors evaluated job performance on a rating scale that ranged from 1-5. The rating scales covered speed and accuracy for nine office skills. The scales also incorporated levels of skill in three areas (when a rating scale was relevant to the job). The nine office skills were: (1) text from hard copy, (2) text from machine dictation, (3) charts/tables/statistics from hard copy, (4) spreadsheet skill, (5) database management skill, (6) data entry skill: numeric, (7) data entry skill: alpha-numeric, (8) ten-key skill, and (9) shorthand/speed writing and transcription skill.

Skills (1) text from hard copy, (2) text from machine dictation, and (3) charts/tables/statistics from hard copy were grouped for a word-processing level of skill rating. Level I included setting tabs, margins, and justification to format documents; using common function keys, such as bold, underline, and center; making simple edits by using delete and insert keys; typing information on pre-printed forms; and naming, saving, printing, and retrieving documents. Level II included setting up, editing, copying, and moving columns; using headers and footers; creating templates and boilerplate formats; creating forms; merging form letters and forms with variable data; creating and printing labels; using DOS commands; using various sizes and styles of lettering; and archiving. Level III included creating and using macros; using graphics; converting documents to ASCII; using math functions; creating a dictionary for the system; and linking spreadsheets or database system information with word-processing documents.

Definitions were provided to the supervisors for each of the skills and scales. Further, supervisors were instructed to only provide input where they had first-hand knowledge.

More than 100 supervisors gave ratings for 292 incumbents during the workshops. Of the 292 incumbents who received ratings, 33 received multiple ratings or ratings from more than one supervisor. For analysis purposes, the multiple ratings were averaged.

Data on the Experimental Tests

The 292 incumbents who received ratings by their supervisors were invited to participate in parts of the experimental test battery. Since supervisors only rated incumbents on skills that were relevant in their situation, and when they had first-hand knowledge of the work behaviors, not all of the 292 incumbents received ratings on all of the skills. Since involvement in the study was voluntary, not all of the 292 incumbents who had received ratings took all of the experimental tests. Testing was conducted over a seven-week period using six PC's. Of the 292 incumbents with ratings, 110 actually took one or more of the tests in the experimental test battery. Of the 110 incumbents taking the tests, 75 took the OPAC Keyboarding Test, 68 took the OPAC Editing/Formatting from Rough Draft Test, and 50 took the OPAC Language Arts 1 Test. A variety of conditions dictated which tests were administered to each incumbent, including the duties the incumbent performed, amount of time the incumbent could spend taking the experimental tests, software and hardware the incumbent used on the job, software and hardware available for testing at the time, sample already obtained in the study, etc. Because of these conditions, the samples varied from test to test.

Job Analysis and Test Evaluation
Some of the incumbents who took the experimental tests also evaluated the tests and provided data as subject matter experts (SME’s). After taking an experimental test, the incumbent was asked to answer (as a subject matter expert) a content validity survey form for that test. If the subject matter expert stated that some level of the skill, which the test measured, was a necessary prerequisite for successful performance of a critical or important job duty, then several other questions were asked. These additional questions asked for a description of the critical or important duties which required use of the skill, then asked for ratings of the degree of importance of that skill. Additional questions subject matter experts answered dealt with the level of the skill which resulted in better performance, if the test was a representative sample of the skill, if the test required more skill from the test taker than was required on the job, if the skill could be learned in a brief orientation, and if the work product of the test closely resembled a work product produced on the job. To obtain information for a job-related cutoff, subject matter experts were given their score and then asked to provide their opinion of the minimum score necessary to pass minimally qualified applicants (following some of the basics of the Angoff model). (See Angoff 1971.)

**Content Validity Results**

If a test product results in adverse impact against a protected group (i.e., one sex, race, or ethnic origin group scores disproportionately lower than another group on the test), the Uniform Guidelines specifically allow content validity as a method of showing business necessity for the test. (See Uniform Guidelines Sections II, 5A, and 14C.) According to the Uniform Guidelines, content validity is:

Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job. (See Uniform Guidelines Section 16D.)

In **Contreras v. City of Los Angeles**, five of seven subject matter experts had to agree on decisions dealing with job relatedness. This standard was accepted by the Court. (See Contreras 1981). In **U.S. v. South Carolina**, 50% of the subject matter experts had to agree for test items to be judged job-related. This standard was accepted by the Court. (See South Carolina 1978.) Therefore, in this study, a minimum standard was set for the content validity of a test when 50% of the incumbents agreed on all the questions in the content validity survey. The preferred standard was set at 70%.

Each test passed all the minimum content validity standards. Therefore, each of the three tests was content valid. In addition, with the exception of word-processing being learned in a brief orientation, every test passed the preferred standards for content validity, as can be seen, in the chart below:

<table>
<thead>
<tr>
<th>Percent of SME’s Who Say:</th>
<th>Keyboarding</th>
<th>Word Processing</th>
<th>Language Arts I Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WPM</td>
<td>Total</td>
<td>WPM</td>
</tr>
<tr>
<td>Some level of skill needed</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Skill is distinguishing</td>
<td>73</td>
<td>70</td>
<td>81</td>
</tr>
<tr>
<td>Test is representative sample</td>
<td>77</td>
<td>80</td>
<td>88</td>
</tr>
<tr>
<td>Test does not require more than job</td>
<td>97</td>
<td>97</td>
<td>91</td>
</tr>
<tr>
<td>Skill cannot be learned in 8 hours</td>
<td>76</td>
<td>76</td>
<td>62</td>
</tr>
<tr>
<td>Test resembles job duty</td>
<td>89</td>
<td>91</td>
<td>99</td>
</tr>
</tbody>
</table>

Copyright © 1989-2006 Biddle Consulting Group, Inc.
34
The following cutoffs were agreed to by at least 70% of the subject matter experts:

<table>
<thead>
<tr>
<th>Test</th>
<th>Scale</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyboarding Test</td>
<td>Error Count Scale</td>
<td>5.00</td>
</tr>
<tr>
<td>Keyboarding Test</td>
<td>Speed Words Per Minute Scale</td>
<td>55.00</td>
</tr>
<tr>
<td>Keyboarding Test</td>
<td>Gross Key Strokes Scale</td>
<td>1443.00</td>
</tr>
<tr>
<td>Editing/Formatting from Rough Draft Test</td>
<td>Total Score Scale</td>
<td>13.00</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Capitalization Scale</td>
<td>.60</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Possessives Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Number Usage Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Abbreviations Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Punctuation Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Spelling Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Grammar Scale</td>
<td>.70</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Total Score Scale</td>
<td>50.00</td>
</tr>
</tbody>
</table>

**Concurrent Criterion-Related Validity Results**

If a test results in adverse impact against a protected group, the Uniform Guidelines specifically allow concurrent criterion-related validity as a method of showing business necessity for the test. (See Uniform Guidelines Sections II, 5A, and 14B(4).) According to the Uniform Guidelines, criterion-related validity is defined as follows:

Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior. (See Guidelines Section 16F.)

**Concurrent** criterion-related validity usually uses current employees as the sample, obtaining test scores and criteria data (e.g., supervisory ratings) during relatively the same time period. **Predictive** criterion-related validity often uses applicants as the sample, obtaining test scores during one period of time, then waiting to gather criteria data later. This study used concurrent criterion-related validity.

Directional hypotheses were set for each scale of the experimental tests. Statistical significance was set at the one-tailed .05 level, specifying the direction of the relationship. Correlations were calculated only for each of the hypothesized relationships using the Pearson Product Moment formula. Restriction in range was a concern, as all the incumbents had passed the employer’s 5 minute timed test in order to obtain their jobs initially. It was suspected that many of the tests in the experimental test battery would correlate with the local 5 minute timed test. However, because the Equal Employment Opportunity field has not established a clear rule allowing for correcting correlations not quite significant into significance, no corrections were made for any possible indirect restriction in range. All correlations presented are uncorrected.

**OPAC Tests**
Below are the correlations calculated between the OPAC tests and supervisory ratings of job performance hypothesized as having a possible relationship. Correlations are index numbers which show the degree of relationship between the test score and supervisory ratings. Correlations range from 1.0, showing a perfect relationship, to 0.0, showing absolutely no relationship. A -1.0 means a perfect inverse relationship - as one score goes up, the other score goes down. Correlations shown below with the asterisk (*) are statistically significant correlations. This means that the degree of relationship is so strong that the relationship is unlikely to be due to chance and chance alone except maybe 5% of the time or less.

Statistically significant correlations were found between the OPAC tests and the Accuracy Rating, Speed Rating, ratings of Level I of Word-processing Skill, and ratings of Level II Word-processing Skill. Very few ratings were obtained on Level III ratings of Word-processing Skill. Each of the Language Arts Test scales statistically significantly correlated with the Accuracy Rating independently, except for the Possessives Scale. The Possessives Scale was close to statistical significance (.259 obtained and .268 needed). The N shown below refers to the number of incumbents who took the tests and had supervisory ratings used in the correlation calculations.

**Correlations Between OPAC Test Scales and Supervisory Ratings**

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Test</th>
<th>Test Scale</th>
<th>Supervisory Performance Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Abbreviations</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.29</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Capitalization</td>
<td>.40</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Grammar</td>
<td>.42</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Number Usage</td>
<td>.38</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Percent Score</td>
<td>.55</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Possessives</td>
<td>.26</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Punctuation</td>
<td>.44</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Spelling</td>
<td>.44</td>
</tr>
<tr>
<td>OPAC</td>
<td>Language Arts</td>
<td>Total</td>
<td>.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Test</th>
<th>Test Scale</th>
<th>Supervisory Performance Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPAC</td>
<td>Word Processing</td>
<td>% Score</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.25</td>
</tr>
</tbody>
</table>

**Editing/Formatting from a Rough Draft**

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Test</th>
<th>Test Scale</th>
<th>Supervisory Performance Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPAC</td>
<td>Word Processing</td>
<td>% Score</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Test</th>
<th>Test Scale</th>
<th>Supervisory Performance Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPAC</td>
<td>5 Min Typing</td>
<td>Incorrect</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-.12(^8)</td>
</tr>
<tr>
<td>OPAC</td>
<td>5 Min Typing</td>
<td>Strokes</td>
<td>.28(^9)</td>
</tr>
<tr>
<td>OPAC</td>
<td>5 Min Typing</td>
<td>WPM</td>
<td>.28(^{10})</td>
</tr>
</tbody>
</table>

\(^{5}\) \(N = 39\).
\(^{6}\) \(N = 66\).
\(^{7}\) \(N = 54\).
\(^{8}\) \(N = 62\).
\(^{9}\) \(N = 59\).
\(^{10}\) \(N = 62\).
Alternative Procedure Analysis

The Uniform Guidelines state that:

Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. (See Uniform Guidelines Section 3B.)

Three other test batteries were included in the study. Therefore, data was available to evaluate "substantially equally valid" and the relative adverse impact of the four test batteries.

Substantially equally valid can be evaluated using content validity and concurrent criterion-related validity. Using content validity, the four test batteries included in this study had substantially equally valid tests with the exception of one test battery's word-processing test for WordPerfect. Using concurrent criterion-related validity, several of the test batteries were close. Many of the key correlations were compared and found to be not significantly different. However, the OPAC test battery was the only test battery to have statistically significant correlations to all four ratings: Speed Rating, Accuracy Rating, Level I of Word-processing Skill Rating, and Level II of Word-processing Skill Rating. Therefore, under concurrent criterion-related validity, no other test battery was substantially equally valid to the OPAC test battery.

An analysis of adverse impact was nevertheless conducted. Since about half or more of the participants in the samples for the tests were Hispanic, adverse impact analyses were feasible.

The Uniform Guidelines requires an analysis of both statistical significance and practical significance in determining adverse impact. (See Uniform Guidelines Section 4D.) For speed purposes, Cochran's correction to the chi-square was used to best approximate statistical significance at the .05 level. (See Haber 1980.) This is a two sample hypergeometric test. Practical significance needs to be addressed to complete the evaluation of adverse impact. (See Uniform Guidelines Section 4D and Baldus 1980). Practical significance with rate differences involves at least three calculations. Each of these calculations involves the effects of small number changes on other statistics. How many more people need to be added to the disadvantaged group's passing number to: (1) change the statistical significance conclusion, (2) change the 80 Percent Rule of Thumb conclusion, or (3) change the selection rates themselves from being different to being the same or very close to being the same. When 2 or fewer people added to the disadvantaged group can alter the statistical conclusion, the results were found to be not practically significant. (See Waisome 1991). When 3 or fewer people added to the disadvantaged group alters the 80 Percent Rule of Thumb conclusion or adding 4 or fewer people brings the selection rates to being very close to one another (within 2.1%), then the results were found to be not practically significant. (See Contreras 1981). Both of these court case citations are Federal circuit court decisions. (For a more detailed discussion of adverse impact see reference: Biddle 1992.)

Using the statistical significance and practical significance rules described above, adverse impact was found for another test battery's spelling scale and OPAC's Keyboarding test for words per minute and key strokes scales. (Since the time of the study, OPAC's Keyboarding test format has been changed. The new format preserves the test that showed the content and criterion-related validity, but now allows the candidate to take the test in a scrolling mode or from hard copy text.)
Overall Conclusions Considering Validity and Adverse Impact

Using criterion-related validity as the standard for "substantially equally valid for a given purpose" for the Section 3B analysis described in this paper, OPAC was the only test battery (in the experimental test batteries) with tests that correlated statistically significantly to all four of the employer's criteria (Speed Ratings, Accuracy Ratings, Level I of Word-processing Skill, and Level II of Word-processing Skill). Since the OPAC test battery was the only test battery that correlated statistically significantly with all four criteria, the other four test batteries cannot be considered "substantially equally valid for a given purpose." The OPAC test battery was able to correlate "above chance levels" to criteria the other tests did not in this situation.
Content and Criterion-Related Validity Report
for the OPAC® System (1994)
A study was conducted at a large federal employer with more than 5000 employees to examine the validity of several OPAC® (Office Proficiency Assessment and Certification®) System tests. The OPAC tests were originally developed and content validated by Professional Secretaries International®. The OPAC tests are computer administered and computer scored.

The employer involved in the study was searching for a self-administered, self-scored computerized word-processing test to replace its traditional 5-minute timed typing test. The traditional typing test measured speed and accuracy using IBM Selectric typewriters and required extensive test administrator supervision. The employer also wanted a test that could measure an applicant's skill at creating, formatting, proofing, and editing documents, while also measuring word-processing skills using a word-processing type program. An applicant's speed and accuracy were also important factors for the test to measure. Additionally, the employer wanted to minimize the time necessary for test administration.

More than 20 job classifications needed a selection testing procedure that measured keyboarding speed and accuracy as well as some level of word-processing skill.

**Experimental Test Battery**

OPAC tests of Language Arts 1, Editing/Formatting from Rough Draft, Advanced Editing/Formatting from Rough Draft, and Keyboarding were used as part of an experimental test battery.

The OPAC Language Arts 1 test evaluated in the study was used to measure skills in proofing a document to identify errors in grammar, spelling, punctuation, capitalization, possessiveness, number usage, and abbreviations.

The OPAC Editing/Formatting from Rough Draft test was used to measure skills in operating features and functions of specific word-processing programs.

The OPAC Advanced Editing/Formatting from Rough Draft test was used to measure skills in operating advanced features and functions of specific word-processing programs.

The OPAC Keyboarding test was used to measure an individual's speed and accuracy of typing text on a keyboard.

**Identification of Sample**

Incumbents of 24 secretarial, clerical, and administrative classifications were sent a survey. The survey asked about the use of word-processing equipment and software. Of the 1139 incumbents who were sent surveys, 65.8% responded (749). The responses showed that WordPerfect, Microsoft Word for the IBM and Macintosh, and MultiMate were the word-processing software most frequently used. Of the 749 incumbents who responded, 94.3% (706) indicated that they were using some form of word-processing on the job. About half (378) used more than one word processor, including text editors or desktop publishing. The 5.7% (43) who used no word-processing on the job were no longer included in the study.
Job Performance Ratings

A series of workshops were conducted for those who supervised the survey respondents to obtain ratings of job performance. Supervisors evaluated job performance on a rating scale that ranged from 1-5. The rating scales covered speed and accuracy for nine office skills. The scales also incorporated levels of skill in three areas (when a rating scale was relevant to the job). The nine office skills were: (1) text from hard copy, (2) text from machine dictation, (3) charts/tables/statistics from hard copy, (4) spreadsheet skill, (5) database management skill, (6) data entry skill: numeric, (7) data entry skill: alpha-numeric, (8) ten-key skill, and (9) shorthand/speed writing and transcription skill.

Skills (1) text from hard copy, (2) text from machine dictation, and (3) charts/tables/statistics from hard copy were grouped for a word-processing level of skill rating. Level I included setting tabs, margins, and justification to format documents using common function keys such as bold, underline, and center, making simple edits by using delete and insert keys, typing information on pre-printed forms, and naming, saving, printing, and retrieving documents. Level II included setting up, editing, copying, and moving columns, using headers and footers, creating templates and boilerplate formats, creating forms, merging form letters and forms with variable data, creating and printing labels, using DOS commands, using various sizes and styles of lettering, and archiving. Level III included creating and using macros, using graphics, converting documents to ASCII, using math functions, creating a dictionary for the system, and linking spreadsheets or database system information with word-processing documents.

Definitions were provided to the supervisors for each of the skills and scales. Further, supervisors were instructed to only provide input where they had first-hand knowledge.

More than 100 supervisors gave ratings for 292 incumbents during the workshops. Of the 292 incumbents who received ratings, 33 received multiple ratings or ratings from more than one supervisor. For analysis purposes, the multiple ratings were averaged.

Data on the Experimental Tests

The 292 incumbents who received ratings by their supervisors were invited to participate in parts of the experimental test battery. Since supervisors only rated incumbents on skills that were relevant in their situation, and when they had first-hand knowledge of the work behaviors, not all of the 292 incumbents received ratings on all of the skills. Since involvement in the study was voluntary, not all of the 292 incumbents who had received ratings took all of the experimental tests. Testing was conducted over a seven-week period using six PC's. Of the 292 incumbents with ratings, 110 actually took one or more of the tests in the experimental test battery. Of the 110 incumbents taking the tests, 75 took the OPAC Keyboarding Test, 68 took the OPAC Editing/Formatting from Rough Draft Test, and 50 took the OPAC Language Arts 1 Test. A variety of conditions dictated which tests were administered to each incumbent, including the duties the incumbent performed, amount of time the incumbent could spend taking the experimental tests, software and hardware the incumbent used on the job, software and hardware available for testing at the time, sample already obtained in the study, etc. Because of these conditions, the samples varied from test to test.
Job Analysis and Test Evaluation

Some of the incumbents who took the experimental tests also evaluated the tests and provided data as subject-matter experts (SMEs). After taking an experimental test, the incumbent was asked to answer (as a subject-matter expert) a content validity survey form for that test. If the subject-matter expert stated that some level of the skill, which the test measured, was a necessary prerequisite for successful performance of a critical or important job duty, then several other questions were asked. These additional questions asked for a description of the critical or important duties which required use of the skill, then asked for ratings of the degree of importance of that skill. Additional questions subject-matter experts answered dealt with the level of the skill which resulted in better performance, if the test was a representative sample of the skill, if the test required more skill from the test taker than was required on the job, if the skill could be learned in a brief orientation, and if the work product of the test closely resembled a work product produced on the job. To obtain information for a job-related cutoff, subject-matter experts were given their score and then asked to provide their opinion of the minimum score necessary to pass minimally qualified applicants following some of the basics of the Angoff model. (See Angoff 1971.)

Content Validity Results

If a test product results in adverse impact against a protected group (i.e., one sex, race, or ethnic origin group scores disproportionately lower than another group on the test), the Uniform Guidelines specifically allow content validity as a method of showing business necessity for the test. (See Uniform Guidelines Sections II, 5A, and 14C.) According to the Uniform Guidelines, content validity is:

Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job. (See Uniform Guidelines Section 16D.)

In Contreras v. City of Los Angeles, five of seven subject-matter experts had to agree on decisions dealing with job relatedness. This standard was accepted by the Court. (See Contreras 1981.) In U.S. v. South Carolina, 50% of the subject-matter experts had to agree for test items to be judged job-related. This standard was accepted by the Court. (See South Carolina 1978.) Therefore, in this study, a minimum standard was set for the content validity of a test when 50% of the incumbents agreed on all the questions in the content validity survey. The preferred standard was set at 70%.

Each test passed all the minimum content validity standards. Therefore, each of the three tests was content valid. In addition, with the exception of word-processing being learned in a brief orientation, every test passed the preferred standards for content validity, as can be seen in the chart below:
(Note: the following cutoffs were agreed to by at least 70% of the subject-matter experts.)

<table>
<thead>
<tr>
<th>Test</th>
<th>Scale</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyboarding Test</td>
<td>Error Count Scale</td>
<td>5.00</td>
</tr>
<tr>
<td>Keyboarding Test Speed</td>
<td>Words Per Minute Scale</td>
<td>55.00</td>
</tr>
<tr>
<td>Keyboarding Test</td>
<td>Gross Key Strokes Scale</td>
<td>1443.00</td>
</tr>
<tr>
<td>Editing/Formatting from Rough Draft Test</td>
<td>Total Score Scale</td>
<td>13.00</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Capitalization Scale</td>
<td>.60</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Possessives Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Number Usage Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Abbreviations Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Punctuation Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Spelling Scale</td>
<td>.50</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Grammar Scale</td>
<td>.70</td>
</tr>
<tr>
<td>Language Arts 1 Test</td>
<td>Total Score Scale</td>
<td>50.00</td>
</tr>
</tbody>
</table>

**Concurrent Criterion-Related Validity Results**

If a test results in adverse impact against a protected group, the Uniform Guidelines specifically allow concurrent criterion-related validity as a method of showing business necessity for the test. (See Uniform Guidelines Sections II, 5A, and 14B(4). According to the Uniform Guidelines, criterion-related validity is defined as follows:

Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior. (Guidelines Section 16F.)

Concurrent criterion-related validity usually uses current employees as the sample, obtaining test scores and criteria data (e.g., supervisory ratings) during relatively the same time period. Predictive criterion-related validity often uses applicants as the sample, obtaining test scores during one period of time, then waiting to gather criteria data later. This study used concurrent criterion-related validity.

Directional hypotheses were set for each scale of the experimental tests. Statistical significance was set at the one-tailed .05 level, specifying the direction of the relationship. Correlations were calculated only for each of the hypothesized relationships using the Pearson Product Moment formula. Restriction in range was a concern, as all the incumbents had passed the employer’s 5 minute timed test in order to obtain their jobs initially. It was suspected that many of the tests in the experimental test battery would correlate with the local 5-minute timed test. However, because the Equal Employment Opportunity field has not established a clear rule allowing for correcting correlations not quite significant into significance, no corrections were made for any possible indirect restriction in range. All correlations presented are uncorrected.

**Criterion-Related Validity Correlations**

Below are the correlations calculated between the OPAC tests and supervisory ratings of job performance hypothesized as having a possible relationship. Correlations are index numbers that show the degree of relationship between the test score and supervisory ratings.
Correlations range from 1.0, showing a perfect relationship, to 0.0, showing absolutely no relationship. A -1.0 means a perfect inverse relationship—as one score goes up, the other score goes down. **Correlations shown above the horizontal lines on the following charts are statistically significant correlations.** This means that the degree of relationship is so strong that the relationship is unlikely to be due to chance and chance alone except maybe 5% of the time or less. Statistically significant correlations were found between the OPAC tests and the Accuracy Rating, Speed Rating, ratings of Level I of Word-processing Skill, and ratings of Level II Word-processing Skill. Very few ratings were obtained on Level III ratings of Word-processing Skill. Each of the Language Arts Test scales statistically significantly correlated with the Accuracy Rating independently, except for the Possessives Scale. The Possessives Scale was close to statistical significance (.259 obtained and .268 needed). The "n" shown below refers to the number of incumbents who took the tests and had supervisory ratings used in the correlation calculations.

**Correlations Between OPAC Test Scales and Supervisory Ratings**

**Language Arts 1 Test (n=39):**

![Language Arts Test Scale Validity](chart.png)

- Abbreviations: 0.29
- Capitalization: 0.40
- Grammar: 0.42
- Number Usage: 0.38
- Possessives: 0.26
- Punctuation: 0.44
- Spelling: 0.44
- OVERALL: 0.65

*0.268 needed for validity*
**Editing/Formatting from Rough Draft:** (Level I n=66; Level II n=54)

**Wordprocessing Levels I and II**

- **Level I**
  - Correlation: 0.25
  - .210 needed for validity (Level 1)

- **Level II**
  - Correlation: 0.25
  - .230 needed for validity (Level 2)

**Word Processing Level**

---

**Keyboarding (Keystrokes n=59; Words Per Minute n=62)**

**Keyboarding Test**

- **Keystrokes**
  - Correlation: 0.3
  - .268 needed for validity

- **Words Per Minute**
  - Correlation: 0.3
  - .252 needed for validity
Alternative Procedure Analysis

The Uniform Guidelines state that:

Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. (See Uniform Guidelines Section B.)

Three other test batteries were included in the study. Therefore, data was available to evaluate "substantially equally valid" and the relative adverse impact of the four test batteries.

Substantially equally valid can be evaluated using content validity and concurrent criterion-related validity. Using content validity, the four test batteries included in this study had substantially equally valid tests with the exception of one test battery's word-processing test for WordPerfect. Using concurrent criterion-related validity, several of the test batteries were close. Many of the key correlations were compared and found to be not significantly different. However, the OPAC test battery was the only test battery to have statistically significant correlations to all four ratings: Speed Rating, Accuracy Rating, Level I of Word-processing Skill Rating, and Level II of Word-processing Skill Rating. Therefore, under concurrent criterion-related validity, no other test battery was substantially equally valid to the OPAC test battery.

An analysis of adverse impact was nevertheless conducted. Since about half or more of the participants in the samples for the tests were Hispanic, adverse impact analyses were feasible.

The Uniform Guidelines requires an analysis of both statistical significance and practical significance in determining adverse impact. (See Uniform Guidelines Section 4D.) For speed purposes, Cochran's correction to the chi-square was used to best approximate statistical significance at the .05 level. (See Haber 1980.) This is a two-sample hypergeometric test. Practical significance needs to be addressed to complete the evaluation of adverse impact. (See Uniform Guidelines Section 4D and Baldus 1980). Practical significance with rate differences involves at least three calculations. Each of these calculations involves the effects of small number changes on other statistics. How many more people need to be added to the disadvantaged group's passing number to: (1) change the statistical significance conclusion, (2) change the 80 Percent Rule of Thumb conclusion, or (3) change the selection rates themselves from being different to being the same or very close to being the same. When 2 or fewer people added to the disadvantaged group can alter the statistical conclusion, the results were found to be not practically significant. (See Waisome 1991.) When 3 or fewer people added to the disadvantaged group alters the 80 Percent Rule of Thumb conclusion or adding 4 or fewer people brings the selection rates to being very close to one another (within 2.1%), then the results were found to be not practically significant. (See Contreras 1981.) Both of these court case citations are Federal circuit court decisions. (For a more detailed discussion of adverse impact see reference: Biddle 1992.)

Using the statistical significance and practical significance rules described above, adverse impact was found for another test battery's spelling scale and OPAC's Keyboarding test for words per minute and keystrokes scales. (Since the time of the study, OPAC's Keyboarding test format has been changed. The new format preserves the test that showed the content
and criterion-related validity, but now allows the candidate to take the test in a scrolling mode or from hard copy text.)

**Overall Conclusions Considering Validity and Adverse Impact**

Using criterion-related validity as the standard for "substantially equally valid for a given purpose" for the Section b analysis described in this paper, the OPAC System was the only test battery (in the experimental test batteries) with tests that correlated statistically significantly to all four of the employer's criteria (Speed Ratings, Accuracy Ratings, Level I of Word-processing Skill, and Level II of Word-processing Skill). Since the OPAC test battery was the only test battery that correlated statistically significantly with all four criteria, the other four test batteries cannot be considered "substantially equally valid for a given purpose." The OPAC test battery was able to correlate "above chance levels" to criteria the other tests did not in this situation.
Content Validity Report for OPAC® Module Four

(March 1997)
OPAC Validity Report: Module 4

Test Description

Biddle Consulting Group, Inc. recently developed a fourth module for the Office Proficiency Assessment and Certification® (OPAC®) System. This module was entitled “10-Key/Data Entry.”

In addition to a 10-Key Test, three different tests were included for the evaluation of data entry skills: the Vendor Test, the Inventory Test, and the Invoice Test. Three data entry test are included within this module to allow an employer to choose the tests that are most appropriate for the job in question. In order to ensure the closest match between the job content and the test materials it is recommended (within Biddle Consulting Group’s OPAC manual) that employers evaluate the content of the tests, the format of the tests, and the percentage of alpha and numeric keystrokes within each document. The goal is for employers to utilize the tests which are closest in content and format to the types of data prospective employees will be expected to enter on the job. Although these three types of tests cannot possibly replicate all types of materials that applicants might come into contact with on the job, they are designed to simulate the format of the most commonly utilized data entry designs. Consequently, an applicant’s performance on these tests will enable an employer to evaluate the individuals general data entry capability.

Biddle Consulting Group recommended that employers evaluate each test before deciding which ones would best serve the their business needs and be the most job-related.

The following information should aid employers in their decision of which test(s) are most appropriate for the job classification under consideration. The tests were presented (as they are within the program) by level of difficulty with the last test having the highest difficulty level.

Vendor Test

These test forms are designed to simulate typical vendor entry sheets. The content of the sheets includes a vendor number, company name, company address, and contact information. These tests are the least difficult of the three data entry tests due to their high percentage of alpha strokes and the general field separation of alphabetic and numeric content (i.e. most fields are either alpha or numeric, except for the address field). The average breakdown of each Vendor test includes 75% alphabetic entry and 25% numeric entry. If individuals hired will be expected to enter fields of information similar to those included on these test forms, these data entry tests may be appropriate to use in a selection process.

Inventory Test

These test forms are designed to simulate typical inventory sheets. The content of the sheets includes item information and vendor number. These tests are more difficult than the Vendor Tests for several reasons. First, the Inventory Tests have a higher numeric content. Second, several fields are not stroke specific. That is, they are a mix of both alphabetic and numeric key strokes, causing more transitions between these two areas on the keyboard. The average breakdown of each Inventory test includes 64% alpha entry and 36% numeric entry. If individuals hired will be expected to enter fields of information similar to those
included on these test forms, these data entry tests may be appropriate to use in your selection process.

**Invoice Test**

These test forms are designed to simulate typical invoice sheets. The content of the sheets includes order number, representative number, date, destination information, and product information. These tests are the most difficult of the data entry tests for two main reasons. First, these tests utilize the highest percentage of numeric keystrokes. Second, although the fields within these tests are generally separated as to alpha and numeric content (most fields are either alpha or numeric, except for the address field), these test forms contain the largest number of data fields. The average breakdown of each Invoice test includes 37% alpha entry and 63% numeric entry. If individuals hired will be expected to enter fields of information similar to those included on the test forms, these data entry tests may be appropriate to use in your selection process.

If you are hiring for a position that embodies different types of data entry, several tests can be given to a candidate to obtain all relevant skill information. It is important to remember, however, that each test varies in content and difficulty level. Candidates will not score the same on each test! It is this variance in difficulty level that led to different certification standards for each test form (see below). The more difficult test will result in a lower SPH score.

<table>
<thead>
<tr>
<th>Test</th>
<th>% Alpha</th>
<th>% Numeric</th>
<th>Certification Level*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-Key</td>
<td>0%</td>
<td>100%</td>
<td>8000 SPH – 95% Accuracy</td>
</tr>
<tr>
<td>Vendor</td>
<td>75%</td>
<td>25%</td>
<td>6200 SPH – 95% Accuracy</td>
</tr>
<tr>
<td>Inventory</td>
<td>66%</td>
<td>34%</td>
<td>5600 SPH – 95% Accuracy</td>
</tr>
<tr>
<td>Invoice</td>
<td>37%</td>
<td>63%</td>
<td>5200 SPH – 95% Accuracy</td>
</tr>
</tbody>
</table>

* These certification levels are not based on national norms. These are preliminary standards, which will be re-evaluated upon further study.

**Review by Biddle & Associates, Inc./Biddle Consulting Group, Inc.**

Module 4 was designed with three test versions within each component. For example, the 10-Key component has a Test Version 1, Test Version 2, and Test Version 3. Each test within module four was reviewed by 13 permanent employees of Biddle & Associates, and nine temporary employees (a total of 22 initial in-house reviews). All 22 individuals evaluated the 12 tests included in Module 4. This initial review included an analysis of the instruction screens, ease of understanding and use of the tests, the testing documents, and the candidate manual. Based on comments from the in-house reviews, improvements and modifications were made to all aspect of the Module 4 program and test forms.

Additional modifications were made to the testing documents after difficulty analyses were performed on all testing materials. Difficulty of the materials was determined by the alphabetic/numeric ratio within and between documents. Alphabetic and numeric characters were described as follows:

**Alphabetic Characters**: For the purposes of the difficulty calculations, an alphabetic character included any character that was not a number. Therefore, alphabetic characters included letters, blank spaces, and symbols (such as &, $,
Punctuation marks were also considered alphabetic due to the fact that they are incorporated within the alphabetic keys and they are no more difficult to type (on average) than the letters on the keyboard. Symbols were included within this category due to the fact that there were so few symbols on the testing documents that they did not merit their own category.

**Numeric Characters**: For the purposes of the difficulty calculations, a numeric character included any number. Decimal points were not counted when they were part of a monetary figure (e.g., 19.95, 29.95).

The symbols that were hard-coded on the screen were not counted within the documents, i.e., they did not contribute to the total key stroke calculations. These included the hyphens (-) within the phone number field (555-555-5555) and the back slashes (/) in the date field (09/09/99).

All tests within each component of Module 4 were evaluated. Each test within a component (e.g., Data Entry 1: Vendor) was modified to ensure that they contained approximately the same percentages of alphabetic and numeric characters. The tests within each component are not statistically different in regards to the alphabetic/numeric ratio. In addition, each test was divided into four quadrants, with the alphabetic/numeric ratio of each quadrant compared to ensure that the difficulty level between different sections of the same test were not statistically different.

**Review by Subject-Matter Experts**

After the in-house (or alpha) review of Module 4, Biddle & Associates conducted an evaluation by individuals outside the company (a beta review). This group included 73 individuals from MTI Business College and Heald Business College in Sacramento, California. Based on comments from the beta review, a number of improvements and modifications were made to the testing program and documents.

Subject-matter experts included individuals from a variety of ethnicities. Participants were predominately female, as data-entry positions generally have an over-utilization of females.

**Development of Certification Levels**

Data from the subject-matter experts from MTI and Heald were also utilized to develop recommended certification levels for the four new sets of tests within Module 4.

The certification levels for the Data Entry tests were developed with a test/test correlation model which utilized a keyboarding score of 45 wpm to predict Data Entry test scores (see Table 1 on the following page). The certification levels for each Data Entry test are different due to the varying alphabetic/numeric content, field lengths, and number of fields per record. Since both speed and accuracy are critical to employers for data entry applications, both were included as certification criteria.

The certification standards for the 10-Key tests were developed by analyzing industry standards for jobs requiring some level of 10-key data entry (many employers require at least 10,000 SPH, but 8,000 SPH is widely accepted as a minimum level for 10-Key speed) and an analysis of beta test scores.
Recommended Certification Levels for Three Data Entry Tests - Biddle & Associates, Inc./Biddle Consulting Group, Inc.

**Data Entry 1 – Vendor**

<table>
<thead>
<tr>
<th>WPM</th>
<th>Predicted Score</th>
<th>Lowest Expected Score</th>
<th>Highest Expected Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>5611</td>
<td>3265</td>
<td>7957</td>
</tr>
<tr>
<td>45</td>
<td>6175</td>
<td>3829</td>
<td>8521</td>
</tr>
<tr>
<td>50</td>
<td>6739</td>
<td>4393</td>
<td>9085</td>
</tr>
<tr>
<td>55</td>
<td>7304</td>
<td>4957</td>
<td>9649</td>
</tr>
<tr>
<td>60</td>
<td>7867</td>
<td>5521</td>
<td>10213</td>
</tr>
</tbody>
</table>

Pearson R = 0.69  
Average Errors = 16  
Recommended Certification Level: 6200 SPH and 95% Accuracy Rate

**Data Entry 2 – Inventory**

<table>
<thead>
<tr>
<th>WPM</th>
<th>Predicted Score</th>
<th>Lowest Expected Score</th>
<th>Highest Expected Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>5245</td>
<td>3460</td>
<td>7029</td>
</tr>
<tr>
<td>45</td>
<td>5610</td>
<td>3826</td>
<td>7395</td>
</tr>
<tr>
<td>50</td>
<td>5975</td>
<td>4191</td>
<td>7760</td>
</tr>
<tr>
<td>55</td>
<td>6341</td>
<td>4556</td>
<td>8125</td>
</tr>
<tr>
<td>60</td>
<td>6706</td>
<td>4922</td>
<td>8491</td>
</tr>
</tbody>
</table>

Pearson R = 0.64  
Average Errors = 15  
Recommended Certification Level: 5600 SPH and 95% Accuracy Rate

**Data Entry 3 – Invoice**

<table>
<thead>
<tr>
<th>WPM</th>
<th>Predicted Score</th>
<th>Lowest Expected Score</th>
<th>Highest Expected Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>4701</td>
<td>1774</td>
<td>7627</td>
</tr>
<tr>
<td>45</td>
<td>5222</td>
<td>2296</td>
<td>8149</td>
</tr>
<tr>
<td>50</td>
<td>5744</td>
<td>2817</td>
<td>8670</td>
</tr>
<tr>
<td>55</td>
<td>6265</td>
<td>3339</td>
<td>9192</td>
</tr>
<tr>
<td>60</td>
<td>6787</td>
<td>3860</td>
<td>9713</td>
</tr>
</tbody>
</table>

Pearson R = 0.57  
Average Errors = 12  
Recommended Certification Level: 5200 SPH and 95% Accuracy Rate

* Scores based on a 95% Confidence Interval.
Accuracy and Completeness

After all modifications were made, all testing documents were entered into the program and checked for 100% accuracy to the key by at least two individuals.
Validation Report for the Medical and Legal Terminology Tests (August 1997)

THE OPAC® SYSTEM version 5.0, Module 5
Introduction

This report contains information regarding the development and content validation of the medical and legal terminology tests within the OPAC® System. The medical and legal terminology tests can be used for the employment, education, or certification of medical assistants and legal assistants or legal secretaries.

Both the medical and legal terminology tests were designed based on content validity standards outlined by the Uniform Guidelines on Employee Selection Procedures, section 14(C).11 The Uniform Guidelines on Employee Selection Procedures provide a single set of principles designed to assist employers, labor organizations, employment agencies, licensing or certification boards comply with Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. These guidelines are a framework for the proper use of tests and other selection procedures.

This report is structured according to the following sub-topics on reporting content validation studies stipulated in section 15(C) of the Uniform Guidelines on Employee Selection Procedures:

1. User(s), locations(s), and date(s) of the study
2. Problem and setting
3. Job analysis
4. Selection procedure and its content
5. Relationship between the selection procedure and the job
6. Alternative procedures investigated
7. Uses and applications
8. Contact person
9. Accuracy and completeness

Although separate validation studies were conducted for the medical and legal terminology tests, these studies will be referred to as one in this report for efficiency and uniformity.

User(s), Locations(s), and Date(s) of the Study

The validation study for the medical and legal terminology tests was completed in July 1997, at Biddle & Associates, Inc./Biddle Consulting Group, Inc., located in Sacramento, California. Medical Assistants were selected for participation in this study from one of the nation's largest Health Maintenance Organizations located at one of its sites in Sacramento, California. Legal Assistants and Legal Secretaries were selected for participation in this study from four large, full service law firms also located in Sacramento, California.

Problem and Setting

The purpose of this validation study was to determine if the knowledge-based terminology tests are a representative sample of the body of learned information that is used, and is a

11 The Uniform Guidelines on Employee Selection Procedures were adopted in 1978 by the Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and the Department of Justice.
necessary prerequisite for the successful job performance of an Entry-level medical assistant and entry-level legal assistant/secretary.

This validation study is predicated upon several important factors. Two industry experts from the medical assistant profession and two from the legal assistant/secretarial field were selected to provide terms and write test items for the medical and legal terminology tests, respectively. (See Appendix for these experts qualifications). These experts were given content validity-based criteria for writing test items. See Appendix for Industry Expert criteria for selection of terms and writing of test items)

The industry experts from the medical assistant field provided two separate lists consisting of two hundred (200) terms each for a combined total of four hundred (400) terms. Both lists were compared and discussed by the experts. Terms that were on both experts’ list were automatically selected for test item writing. Terms that were not on both lists were discussed by the experts and either discarded or selected for item writing. This process resulted in the selection of two hundred (200) medical terms test item construction. A preliminary medical terminology test was designed consisting of the selected terms.

The same process as stated above went into the design of the preliminary legal terminology test. The only difference is that one hundred and fifty terms (150) were selected for the preliminary legal terminology test design. Thus, the preliminary test for legal terminology contains one-hundred (150) test items.

Thirty-nine (39) Medical Assistants and twenty-five (25) Legal Assistants/Secretaries were selected as Subject Matter Experts. All Subject Matter Experts were required to currently hold the job title for the target positions and have at least one-year experience. A majority of the Subject Matter Experts selected had several years’ experience individually in the above classifications.

The thirty-nine (39) Medical Assistants and twenty-five (25) Legal Assistants and Legal Secretaries took the preliminary medical and legal terminology tests, respectively.

The tests and surveys were collected and processed. The tests along with Scantron answer sheets were loaded into the Test, Scoring & Analysis System program which is a testing software package developed by Biddle Consulting Group, Inc. After this process was complete, a minimum cutoff score (pass/fail score) was set based upon calculations using the modified Angoff method. Survey forms were also designed to assess the content validity of each item on both the medical and legal terminology tests. The survey forms are titled Test Response Survey. Twenty-two (22) of the Medical Assistant Subject Matter Experts and (25) Legal Assistants and Legal Secretaries completed the Test Response Survey forms.

Survey forms were also designed to assess the content validity of each item on both the medical and legal terminology tests. The survey forms are titled Test Response Survey. Twenty-two (22) of the Medical Assistant Subject Matter Experts and (25) Legal Assistants and Legal Secretaries completed the Test Response Survey forms.

Test Scoring & Analysis System is a comprehensive and proven tool for developing, administering, scoring, and tracking objectively scored tests. This system also has programs that provide data on test-item analysis, test distribution results, and statistical cutoff-score analysis.

The modified Angoff method involves the setting of a job-related minimum cutoff score for test that has been approved by the United States Supreme Court in the case U.S. v. South Carolina, 15 EPD 7, 920, 445 F. Supp 1094 (DC S.Ct.1977) and 15 EPD 8027434 US 1026 (1978).
Results from the surveys were used to conduct validation analysis. Several test items were eliminated during this process. After the validation analysis, the medical and legal terminology tests contained one hundred and sixty-two (162) and seventy-five (75) test items, respectively.

There are limiting factors in the size and scope of this validation study that may affect the validity of these tests for general use. The test items were constructed based on the opinions and experience of industry experts from one city. The subject matter experts who took the test and responded to the survey forms were also selected from one city. There were no studies conducted using a control group to show that the test distinguishes statistically between candidates who have the prerequisite knowledge to perform the task associated with the specified positions and those who do not.

In addition, there were no validity analyses for the rank ordering of test scores above the minimum cutoff scores. Therefore, the recommended minimum cutoff score for each test is valid only for pass/fail purposes. In other words, the test distinguishes only between passing or failing scores and does not provide a basis for ranking scores above the cutoff score.

Given these limiting factors, Biddle Consulting Group, Inc., recommends that users convene a group of their own subject-matter experts to determine if the test is valid for their purpose and specific job classifications. The OPAC System has a validation module that is designed for users to conduct their own validity study.

**Job Analysis**

Knowledge of medical or legal terminology was deemed a necessary prerequisite for the performance of medical assistant or legal assistant/secretary classifications based on job descriptions industry experts’ opinions, and surveys completed by subject-matter experts. Thus, the focus of this study centers on the identification of the specific knowledge that is used and a necessary prerequisite for the work behaviors of medical assistants and legal assistants and or legal secretaries.

Knowledge of medical terminology has a direct relationship to the work behavior of a Medical Assistant because it is important and necessary for communication, record maintenance, and treatment of patients. Knowledge of legal terminology is related to the work behavior of Legal Assistant/Secretaries because it is important and necessary for communication, research, and preparation of legal documents.

Moreover, the focus of this validation study is based on the analysis of medical and legal terminology test items that meet the standards of content validity for knowledge-based selection procedures outlined by the Uniform Guidelines on Employee Selection Procedures, section 14(C) 4, which hold:

For any [test] measuring a knowledge...the user should show that (a) the [test] measures and is a representative sample of that knowledge...and (b) that knowledge...is used in and is a necessary prerequisite to performance of critical or important work behavior(s).

Subject-matter experts for the chosen classifications were given criteria to analyze each test item using the Test Survey Response form which address the above guidelines. This survey was constructed based on models presented in the Guidelines Oriented Job Analysis (GOJA) offered by Biddle (1996). The GOJA® method has been supported in numerous court cases.
for a variety of jobs. Essential components of the survey used by the subject-matter experts follow:

<table>
<thead>
<tr>
<th>Categories</th>
<th>Ratings with Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Ans.</td>
<td>Write “yes” or “no” to indicate whether the answer provide is correct.</td>
</tr>
<tr>
<td>Frequency Rating</td>
<td>Write the letter(s) that indicate the frequency (how often) the term is used on the job (e.g. in correspondence, reading material, conversation).</td>
</tr>
<tr>
<td></td>
<td>D = Daily</td>
</tr>
<tr>
<td></td>
<td>W = Weekly</td>
</tr>
<tr>
<td></td>
<td>BW = Bi-Weekly</td>
</tr>
<tr>
<td></td>
<td>M = Monthly</td>
</tr>
<tr>
<td></td>
<td>BM = Bi-Monthly (every two months)</td>
</tr>
<tr>
<td></td>
<td>Q = Quarterly</td>
</tr>
<tr>
<td></td>
<td>SA = Semi-Annually</td>
</tr>
<tr>
<td></td>
<td>A = Annual</td>
</tr>
<tr>
<td></td>
<td>LA = Less often than once a year</td>
</tr>
<tr>
<td>Importance</td>
<td>Provide one of the following ratings to indicate how important knowledge of the term is to the job.</td>
</tr>
<tr>
<td></td>
<td>1. NOT IMPORTANT: Trivial or minor significance to the performance of the job.</td>
</tr>
<tr>
<td></td>
<td>2. SOMEWHAT IMPORTANT: Somewhat helpful, useful, and/or meaningful to performance of the job.</td>
</tr>
<tr>
<td></td>
<td>3. IMPORTANT: Helpful, useful, and/or meaningful to the performance of the job.</td>
</tr>
<tr>
<td></td>
<td>4. CRITICAL: Necessary for the performance of the job.</td>
</tr>
<tr>
<td></td>
<td>5. EXTREMELY CRITICAL: Necessary for the performance of the job, but with more extreme consequences</td>
</tr>
<tr>
<td>% Of Qualified Apps.</td>
<td>To your best estimate, indicate the percentage of minimally qualified applicants that would be expected to answer the particular question correctly.</td>
</tr>
<tr>
<td>When Required</td>
<td>Write the letter that indicates when knowledge of this term must be known.</td>
</tr>
<tr>
<td></td>
<td>A. Required at time of hire</td>
</tr>
<tr>
<td></td>
<td>B. Learned on the job</td>
</tr>
</tbody>
</table>

The relationship between the terminology tests and the target positions was established based on the averages of Importance, Frequency, and When Required ratings that all subject-matter experts provided for each test item.

Item difficulty ratings were obtained for each test question along with an overall item difficulty rating for both the medical and legal terminology tests. These ratings were obtained from the Test Scoring & Analysis System software. Item difficulty shows the proportion of subject-matter experts who answered an item correctly.

**Selection Procedure and Contents**
There are two medical terminology tests consisting of eighty (80) multiple choice test items each, labeled Medical Test Form “A” and Medical Test Form “B.” There is one legal terminology test consisting of seventy-five (75) multiple choice test items.

The above tests are part of the 1997 release of the OPAC® System version 5.0. This version of the OPAC System is commercially available and distributed by Biddle Consulting Group, Inc.

Industry experts along with the Product Development Analyst wrote the test items according to item writing criteria. The criteria for Test Item Writing were composed by the Product Development Analyst. These criteria are based on data for writing test items provided by Biddle Consulting Group, Inc., and principles for constructing test items offered by Osterlind (1989).

As stated above, one-hundred-and-sixty-two (162) medical terminology test items were selected for the final item bank. Selection of these items was based on validation criteria applied to averaged results obtained from the Test Survey Response forms. Fifty percent (50%) or more of the Subject-matter experts assigned an importance rating of “3” or greater to each one of the 162 test items. This means that knowledge of each of the 162 medical test items is considered either important, critical, or extremely critical to the performance of the Medical Assistant classification by at least fifty percent, ¹⁴ of the Subject-matter experts surveyed.

Two parallel test forms were created from the 162 test items which represent Medical Terminology Test Form “A” and Medical Terminology Test Form “B.” Both test forms contain eighty (80) test items each. These test forms have the same type of material, with the same level of difficulty, but different test items.

Medical terminology tests forms “A” and “B” each have been determined to measure and represent a sample of the knowledge of medical terminology that is used and is a necessary prerequisite in the job performance of the Entry-level medical assistant classification. This determination is based on (a) the results of the validation study involving the analysis of responses to twenty (22) medical assistant Test Survey Response forms and (b) analysis of the test distribution results.

The Legal Terminology test also has been determined to measure and represent a sample of the knowledge of legal terminology that is used and is a necessary prerequisite to the job performance of an Entry-level legal assistant and or Legal Secretary. This determination is also based on (a) the results of the validation study involving the analysis of responses of twenty-five (25) Legal Assistants and Legal Secretaries Test Survey Response forms and (b) analysis of the test distribution results.

Seventy-five (75) legal terminology test items were selected for the final item bank. Selection of these items is based on validation criteria applied to averaged results obtained from the Test Survey Response forms. Fifty percent (50%) or more of the Subject-Matter Experts assigned an importance rating of “3” or greater to each one of the 75 test items (see Importance ratings above). This means that knowledge of each of the 75 test items is considered either important, critical, or extremely critical to the performance of the Legal

¹⁴ The standard that at least 50 percent of the Subject-Matter Experts need to agree on issues that determine inclusion of an item on a test was approved by the U.S. Supreme Court in the court case U.S. v. South Carolina, 434 US 1026 (1989).
Assistant and or Legal Secretary classification by at least fifty percent of the Subject-matter experts surveyed.

**Relationship between the Selection Procedure and the Job**

The evidence demonstrating that the medical and legal terminology test items are a representative sample of the knowledge used as a part of the work behavior of medical assistants and legal assistants/secretaries was obtained from information on the Test Survey Response reported by subject-matter experts.

Entries from the surveys where compiled into two reports--medical and legal--using a spreadsheet program. The report from the Medical Assistant subject-matter experts surveyed has thirty-one (31) pages and contains twenty thousand (22,000) entries. This report will be referred as the Medical Survey Report. The report from the Legal subject matter experts has twenty (20) pages and contains eighteen thousand plus (18,000+) entries. This report will be referred to as the Legal Survey Report.

Both reports were then imported into a database program, and subset reports were then created from them. The subset reports provide average ratings for each test item which is calculated by category--Correct Ans, Frequency, Importance, and When Required. The subset reports were used to conduct validation analysis of each test-item Test-items were selected or deselected for the final test item bank using the following criteria (which will be referred to as Test-Item Validation Criteria):

1. At least 50 percent of the Subject Matter Experts surveyed agree that the knowledge of the specific test item is required at the time of hire.

2. At least 50 percent of the Subject Matter Experts surveyed rated that knowledge of the specific test item “3” or greater in Importance.

   **AND**

3. At least 50 percent of the Subject Matter Experts surveyed indicated that the term is used annually or more frequently on the job.

As noted above, the United States Supreme Court approved that a fifty percent agreement among subject-matter experts is an acceptable standard for the inclusion of an item on a test (U.S. v South Carolina, 1978).

Micro-reports for both the medical and legal tests were created from the subset reports to show the specific criteria that each selected test item meets.

The two parallel medical test forms that were created from the 162 test items selected are presented in. Again, both test forms contain eighty (80) test items each. These test forms have the same type of material, with the same level of difficulty, but different test items.

Results of the validation study indicate that Medical tests forms “A” and “B” each measures and represents a sample of the knowledge of medical terminology that is used and is a necessary prerequisite in communication, record keeping, and treatment of patients to perform the job of an Entry-level Medical Assistant.

Measures of central tendency, standard deviation, and estimates of reliability were computed using the Test Scoring & Analysis System software for Medical Terminology Tests.
forms "A" & "B." The following test results are based on the Subject Matter Experts’ test scores:

<table>
<thead>
<tr>
<th>Test Form A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Items =</td>
<td>80</td>
</tr>
<tr>
<td>Number of Subjects =</td>
<td>39</td>
</tr>
<tr>
<td>Test Mean =</td>
<td>65.85</td>
</tr>
<tr>
<td>Standard Deviation =</td>
<td>8.639</td>
</tr>
<tr>
<td>Test Reliability =</td>
<td>.8837</td>
</tr>
<tr>
<td>Average Test Difficulty =</td>
<td>.8230</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medical Test Form B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Items =</td>
<td>80</td>
</tr>
<tr>
<td>Number of Subjects =</td>
<td>39</td>
</tr>
<tr>
<td>Test Mean =</td>
<td>62.26</td>
</tr>
<tr>
<td>Standard Deviation =</td>
<td>9.901</td>
</tr>
<tr>
<td>Test Reliability =</td>
<td>.8958</td>
</tr>
<tr>
<td>Average Test Difficulty =</td>
<td>.7782</td>
</tr>
</tbody>
</table>

Results of the validation study indicate that the legal test form measures and represents a sample of the knowledge of legal terminology that is used and is a necessary prerequisite in communication, record keeping, and preparation of legal documents to perform for the job of a Legal Assistant or Legal Secretary.

Measures of central tendency, standard deviation, and estimates of reliability were also computed for the legal terminology test. Some of these results follow:

<table>
<thead>
<tr>
<th>Legal Form</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Items =</td>
<td>75</td>
</tr>
<tr>
<td>Number of Subjects =</td>
<td>25</td>
</tr>
<tr>
<td>Test Mean =</td>
<td>64.52</td>
</tr>
<tr>
<td>Standard Deviation =</td>
<td>8.080</td>
</tr>
<tr>
<td>Test Reliability =</td>
<td>.9025</td>
</tr>
<tr>
<td>Average Test Difficulty =</td>
<td>.8602</td>
</tr>
</tbody>
</table>

Alternative procedures investigated

No alternative test or selection procedure was investigated for this study. Nor were adverse impact analyses conducted. Nevertheless, content validity has been demonstrated for all the tests that justify their use on the grounds of business necessity. The Uniform Guidelines, section II, specifically allow content validity as a method of showing business necessity for the use of a selection procedure (test).

Uses and applications

The medical and legal terminology tests are intended for use in employment, training, education, certification, or other related purposes. As indicated above, these knowledge-
based tests have been shown to represent a sample of the knowledge that is used and is a necessary prerequisite for the successful job performance of an Entry-level Medical Assistant and a Legal Assistant and or Legal Secretary.

These tests were designed to be used primarily as a screening device for hiring, training, education, licensing, certification, or other related purposes. The test scores should be used on a pass/fail basis only. The following cutoff scores (pass or fail scores) are recommended:

<table>
<thead>
<tr>
<th>Test</th>
<th>Pass score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Terminology Form “A”</td>
<td>55</td>
</tr>
<tr>
<td>Medical Terminology Form “B”</td>
<td>53</td>
</tr>
<tr>
<td>Legal Terminology</td>
<td>50</td>
</tr>
</tbody>
</table>

Each test item on all tests is weight one (1.0) or worth one point. A score of fifty-five (55) of a total possible score of eighty (80) is the minimum passing score recommended for Medical Terminology Form “A”. A score of fifty-three (53) of a total possible score of eighty (80) is recommended for Form “B”. Similarly, the minimum passing score recommended for the Legal Terminology Test is fifty (50).

The purpose for setting these cutoff scores is to distinguish between candidates who have demonstrable knowledge of medical or legal terminology that is used and is a necessary prerequisite for successful job performance (for the jobs stated above) and those who do not have this knowledge.

The above cutoff scores were derived from a job-related cutoff setting process called the modified Angoff method. This method involves establishing an overall average level of minimum proficiency using several subject matter experts and then lowering the average rating by one standard error of measurement. The United States Supreme Court has accepted the modified Angoff method for setting job-related cutoff scores for tests (U.S. v. South Carolina, 1978).

All subjects matter experts were required to provide a percentage rating via the Test Survey Response form for each test item based on their opinion of minimum qualified applicants that would be expected to answer the question correctly level. This category is termed Percentage of Qualified Applicants and appears under column four on the Test Survey Response form. Averages were calculated per test item. These averages were loaded into the Job Related Cutoff program of TSA software system. This program calculated an overall average percentage using the averaged score per test item for the tests. The overall average score was then lowered by one standard measure of error (modified Angoff method). The standard measure of error was calculated by TSA and appears as part of the Test Distribution Results. This process resulted in the setting of the cutoff scores, listed above, for each test.

**Contact person**

The person who may be contacted for further information about this validity study is:

James Kuthy, M. A.
Senior Consultant

---

15 The standard error of measurement is designed for interpreting the reliability of test scores. It is used to distinguish between test scores that are statistically different.
Accuracy and completeness

To ensure accuracy and completeness all survey entries were checked and compared. Item difficulty levels were compared to subject matter experts’ minimum qualified applicant ratings. Wherever item difficulty rating was significantly lower than the subject-matter experts’ expected proficiency rating, the subject matter experts’ rating was adjusted to equal the item difficulty rating. This procedure prevents overestimation of ratings, which avoids inflated cutoff scores. The Correct Answer columns were checked for all survey responses. Any test item that did not receive a rating of one hundred percent (100%) agreement regarding its correctness was check thoroughly and adjusted where necessary. Any test item indicating a negative correlations with the key was checked and adjusted (this correlation was provided by Item Analysis program in TSA).
Development Report for OPAC® System 5.0

Legal Keyboarding and Language Arts Tests

October 1998
Disclaimer

Though the research conducted for this report is thorough and complete, it should in no way be construed as a final validation study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the skill(s) being tested. Because this study was conducted at only one employer, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the skills measured by the tests in this report are essential to the specific job environment in which the in-house validation study was conducted.
Abstract

Legal keyboarding and language arts tests were developed to aid in the selection of properly qualified candidates in the legal assistant and legal secretary job classifications. Three alternative versions of each test were developed. The legal keyboarding test was designed to measure the speed and accuracy of applicants typing legal text. The legal language arts test was designed to measure an applicant’s ability to proofread and spot various grammatical errors in documents that legal assistants and secretaries would typically be expected to analyze and proofread. Two legal industry experts assisted in the development of both tests, and 39 subject-matter experts participated in the evaluation of the new tests. One hundred percent of legal subject-matter experts who examined the keyboarding test agreed that the test appropriately measured the skill being assessed, and 100% of subject-matter experts who examined the language arts test also agreed that the test appropriately measured the skills being assessed. Legal subject-matter experts were administered all alternative forms of both tests, and their input established alternate form reliability coefficients for each test. Cutoff scores were also derived, based upon the scores of job incumbent subject-matter experts.
Background

The following is a report describing the development process of the OPAC System legal keyboarding and language arts tests. The reason for developing these tests was twofold. First, a product development decision had been made to orient the OPAC System towards the legal industry, as there is a high need for clerical skills in this field and a perceived high demand for skills testing in legal industry. Second, informal feedback from representatives of the legal industry (solicited mainly from tradeshows conventions and telephone interviews) suggested that legal keyboarding and language arts tests might be the most needed tests for the industry, and thus the most likely tests to develop. Additionally, the OPAC System already contained general versions of these tests, so there was both a product history and test format from which to develop the new instruments.

Early Development

Although an informal perusal into the job occupations of legal assistant and secretary revealed that both keyboarding and language arts skills were important to successful performance in these job classifications, more quantifiable evidence needed to be obtained. To that end, 241 law offices throughout the United States were contacted via facsimile and asked to provide job descriptions for the positions of legal assistant and secretary. Out of the 241 offices contacted, 11 provided complete job descriptions for these positions. All of the received job descriptions indicated that at least some level of minimum competency in the skills of keyboarding and language arts was needed for successful performance in these job classifications. This information provided enough evidence to justify the full development of selection tests measuring keyboarding and language arts skills. Appendix 1 contains all received legal assistant and secretary job descriptions.

Industry Experts

Industry experts were recruited to provide guidance and direction in the test development process. Two industry experts participated in constructing the tests. All industry experts were required to have at least five years of experience in a job classification at or above the level of legal assistant or secretary (the qualifications of these experts are provided in Appendix 2). It was the duty of the industry experts to first provide materials from which to develop the tests, and then to provide feedback and advice on how to develop the tests. Based on the material provided by industry experts, three alternate versions of each test were developed. Once completed, the tests were shown to industry experts, who then evaluated them as to their content and provided recommended changes. The tests were revised and again presented to industry experts for final approval. Industry experts were compensated for their participation in the test development process.

Test Descriptions

The legal keyboarding test was designed to measure typing speed and accuracy specific to legal documents frequently typed by legal assistants and secretaries. Three alternate versions of the test were constructed. Each version had between 640 and 692 words of text. The text material was selected from actual documents that had been used in a several law offices and the tests were similarly formatted to take into account form, content, and layout of the presented text. All tests were constructed to have roughly the same overall level of difficulty. To distinguish it from regular typing tests, the legal keyboarding test contains frequently used legal terminology and other such legal-specific contents. Because of the
frequent technical and numeric information contained in the test, it was thought that skill performance differences between the legal keyboarding test and a non-specific keyboarding test might vary, with test takers performing better on a non-specific test (that does not contain the highly technical information found in the legal keyboarding test). In its final format, the legal keyboarding test will be presented to candidates either on a computer screen, or on a hardcopy printout. Appendix 3 contains all three versions of the test.

The legal language arts tests were designed to measure grammar and proofreading skills. As with the legal keyboarding test, three alternate versions of the test were constructed, and these versions were constructed with the intention of being similar in both structure and difficulty level. The test was designed to simulate actual legal documents, such as a Request for Production or a will. A series of errors were imbedded in the text, the goal for the test taker being to locate and correct these errors. The errors were divided into the classifications of spelling, grammar, punctuation, number usage, possessives, and capitalization. To successfully complete the test, candidates must not only identify the errors (demonstrating proofreading skill), but also have the knowledge to correct the uncovered errors. Each alternate version of the test had between 78 and 82 errors imbedded in the text document, which was between 329 and 356 words long. Error-to-total-word ratios ranged from .23 to .24, which is a similar level found in current OPAC System language arts tests. Appendix 4 contains all three versions of this test.

Testing Site

After construction of the tests was complete, it became necessary to locate a suitable testing site from which to pilot test the new instruments. For the legal keyboarding and language arts tests, a large law office located in Menlo Park, California was selected as the testing site for the new instruments. This test site offered a large pool of subject-matter experts from which to draw, and it also provided subject-matter experts who had some diversity in their particular area of law practice. Subject-matter experts from several fields of law were able to participate in the study.

Method

Participants

Thirty-nine subject-matter experts took part in the beta testing of the legal keyboarding and language arts tests (N = 39). All subject-matter experts were either legal assistants or legal secretaries (or of similar classification) and had at least one year of experience working in that job occupation. The overall mean years of job experience for the subject-matter experts was 9.81 (M = 9.81, SD = 7.13). Subject-matter experts spent approximately one hour taking and evaluating all three versions of both tests. Upon completing the test evaluation, subject-matter experts were thanked for their participation and compensated for their time with gift certificates from a local department store.

Materials

Legal Keyboarding and Language Arts Tests.

Final beta versions of the legal keyboarding and language arts tests were administered to subject-matter experts. The tests were contained in a special beta version of OPAC 5.0 skills testing software that had been installed onto six computers in the law office’s training room.
Candidates were able to open the program by selecting an icon located on the desktop of the computer. Once opened, the computer automatically launched the tests, and candidates completed all three versions of each test.

Validation Survey.

The validation survey was used to evaluate the quality and content validity of each test being examined. The survey was constructed based on a validation report included in OPAC 5.0, and addresses the content validation requirements described in the Uniform Guidelines (1978). Data on each topic was gathered in the survey:

- Whether or not the test measured the skill it was designed to measure
- Whether or not the skill being measured is required at job entry
- The importance of the skill
- The difficulty level of the test
- The subject-matter expert’s score on the test
- The subject-matter expert’s opinion as to what a minimally qualified candidate’s score on the test should be to be considered for employment/promotion

The survey was also designed to capture subject-matter expert demographic information such as name, gender, ethnicity, job title, and years of work experience. All versions of both tests were examined separately, and subject-matter experts completed validation surveys for all versions of each test.

Procedure

A training supervisor at the law office was placed in charge of the test site. Subject-matter experts were tested in groups of five or six during their lunch hour. These testing sessions were staggered out over a one-week period, so as to allow sufficient time for each subject-matter expert to be able to participate. Subject-matter experts were seated at the computer which had the beta version of the OPAC software installed. Once seated, subject-matter experts were given the validation survey, which contained full instructions on how the testing process was to proceed. In order to keep track of their scores on the computer, subject-matter experts entered their social security number when prompted to do so by the computer. The computer then administered each version of the both tests to subject-matter experts, who had five minutes to complete each keyboarding test, and 13 minutes to complete each language arts test. The order in which the tests were presented was randomized so as to lessen any carry-over or practice effects. Between each test, the computer was paused, allowing subject-matter experts to answer validation questions about each of the tests in the survey.

After all six tests were completed, the subject-matter experts were asked to attest that they gave each test their best effort, which they did by checking a box on the last page of the survey that indicated as such. Subject-matter experts were thanked for their time and escorted from the test site.

Results

In order to establish basic content validity for each test, at least 50% of subject-matter experts must agree that proficiency in the skill which the test measures is essential for successful performance of the job being selected for. One hundred percent of subject-matter experts agreed that proficiency in language arts was essential to successful
performance in the job of legal assistant or secretary, and all agreed that keyboarding skills were necessary for successful performance of the job of legal assistant or secretary.

It is also essential to demonstrate that a skill being tested for is required at the time of job entry, and cannot be learned during a brief orientation. To that end, subject-matter experts were asked whether or not keyboarding and language arts skills were required at time of job entry or if they could be learned while on the job. Eighty-eight percent of subject-matter experts agreed that language arts skills were essential at time of job entry, and 82% percent agreed that keyboarding skills were essential at time of job entry.

**Legal Keyboarding**

Each alternate version of the legal keyboarding test was examined to determine mean scores and difficulty levels for each. Mean scores and standard deviations for the legal keyboarding test versions one, two, and three were highly comparable ($M = 63.16$, $SD = 14.79$, $M = 64.08$, $SD = 15.42$, $M = 65.72$, $SD = 14.97$), suggesting that the tests contained similar content and had a similar level of difficulty. The overall mean standard error of measurement was 5.81. In order to determine consistency between the different versions of the test, an alternate form reliability analysis was conducted. The Pearson product-moment correlation coefficient was used to determine the reliability of each version of the test. From this analysis, the following matrix was developed.

**Table 1: Product-moment correlations between each version of the Legal Keyboarding Test.**

<table>
<thead>
<tr>
<th></th>
<th>Legal Keyboarding Version One</th>
<th>Legal Keyboarding Version Two</th>
<th>Legal Keyboarding Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Keyboarding</td>
<td>1.00</td>
<td>.930*</td>
<td>.819*</td>
</tr>
<tr>
<td>Version One</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Keyboarding</td>
<td>.930*</td>
<td>1.00</td>
<td>.799*</td>
</tr>
<tr>
<td>Version Two</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Keyboarding</td>
<td>.819*</td>
<td>.799*</td>
<td>1.00</td>
</tr>
<tr>
<td>Version Three</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.

Based upon the correlations between each version of the test, an overall mean correlation was determined, $R (38) = .85$, $p < .01$. This is a strong reliability coefficient, and indicates consistency between different versions of the test. Subject-matter experts were also asked to rate the difficulty level of the test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts rated the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of $M = 2.09$, $SD = 0.55$, indicating that the tests are set at an appropriate level of difficulty.
Legal Language Arts

As with the legal keyboarding tests, each alternate version of the legal language arts test was examined to determine mean scores and difficulty levels. Mean scores and standard deviations for the legal language arts test versions one, two, and three were consistent ($M = 63.77, SD = 9.21, M = 60.24, SD = 11.19, M = 64.38, SD = 7.90$), meaning that the tests contained similar content and had a similar level of difficulty. Overall, the mean standard error of measurement was 4.50. As with the legal keyboarding tests, a reliability analysis was conducted. The Pearson product-moment correlation coefficient was again used to determine the reliability of each version of the test. From this analysis, the following matrix was constructed.

Table 2: Product-moment correlations between each version of the Legal Language Arts Test.

<table>
<thead>
<tr>
<th></th>
<th>Legal Language Arts Version One</th>
<th>Legal Language Arts Version Two</th>
<th>Legal Language Arts Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Language</td>
<td>.855*</td>
<td>.757*</td>
<td>.855*</td>
</tr>
<tr>
<td>Arts Version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Language</td>
<td>.757*</td>
<td>1.00</td>
<td>.736*</td>
</tr>
<tr>
<td>Arts Version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Language</td>
<td>.855*</td>
<td>.736*</td>
<td>1.00</td>
</tr>
<tr>
<td>Arts Version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.

An overall mean correlation was determined, $R (38) = .78, p < .01$. This is an acceptable reliability coefficient, and indicates consistency between different versions of the test. Subject-matter experts were lastly asked to rate the difficulty level of the language arts test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts rated the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of $M = 2.17, SD = 0.63$, indicating that the tests are set at an appropriate, if slightly high, level of difficulty.

Angoff Scores

To determine the appropriate cutoff score for each test, the modified Angoff method was utilized. The United States Supreme Court (U.S. v. South Carolina) has upheld this method of determining test cutoff scores (Biddle, 1993). Subject-matter experts were asked as to what they believed the score on each test for a minimally qualified applicant should be, which is designed to represent how a minimally qualified job applicant would perform on the test. Subject-matter experts provided these Angoff scores for all versions of each test. Angoff scores were then averaged across alternate versions of each test, yielding a mean Angoff score of 54.20 for the legal keyboarding test, and 56.88 for the legal language arts tests. Based on these Angoff scores, cutoff scores using each test’s standard error of measurement could be derived. The cutoff score for each test was set at one standard error of measurement unit below the test’s mean Angoff. This process led to the following modified Angoff cutoff score for each test.
Table 3: Summary Statistics and Modified Angoff Cutoff Scores for the Legal Keyboarding and Legal Language Arts Tests.

<table>
<thead>
<tr>
<th></th>
<th>Legal Keyboarding</th>
<th>Legal Language Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Angoff Score</strong></td>
<td>54.20</td>
<td>56.88</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>14.97</td>
<td>9.66</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>.85</td>
<td>.78</td>
</tr>
<tr>
<td><strong>Mean Standard Error of Measurement</strong></td>
<td>5.81</td>
<td>4.50</td>
</tr>
<tr>
<td><strong>Modified Angoff Cutoff Score</strong></td>
<td>48</td>
<td>52</td>
</tr>
</tbody>
</table>

Appendix 5 contains full summary statistics for each test, as well as raw candidate scores and feedback from each of the selection tests.

**Performance Differentiation**

Lastly, subject-matter experts were polled to determine how strongly they believed that higher levels of mastery in the skill being assessed distinguished candidates with higher levels of performance in a particular job duty from candidates with lower levels of performance in this job duty. Using a Likert-type scale ranging from 1 to 4 (1 indicating little or no performance differentiation, 2 indicating some performance differentiation, 3 indicating significant performance differentiation, and 4, indicating very significant performance differentiation) subject-matter experts were asked to rate how performance differentiating the skills being assessed by the new tests were. Subject-matter experts gave the legal keyboarding test a mean performance differentiation rating of 2.59, and the legal language arts test a mean performance differentiation rating of 2.75, suggesting that higher levels of these skills may be performance differentiating.

**Job Duty/KSA Linkage**

The Uniform Guidelines (1978) require that tested knowledge, skills, and abilities (KSAs) be linked to established job duties. Responses from subject-matter experts almost universally agreed that keyboarding and language arts skills were essential components of major job duties. Subject-matter experts were asked to list the two most important job duties that link to the tested KSAs, and to rank the importance and frequency of each job duty. Job duties such as “processing of documents,” “transcription,” and “drafting correspondence” were linked to both keyboarding and language arts skills by subject-matter experts. See Appendix 5 for full descriptions. On a Likert-type scale of 1 to 5 (1 being not important, 5 being extremely critical), subject-matter experts rated the overall importance of listed job duties with a mean rating of $M = 3.68$, indicating that the linked job duties were essential to successful job performance. Subject-matter experts also assigned a frequency rating to the listed job duties. Using a Likert-type scale of 1 to 5 (1 indicating daily to weekly performance of the job duty, 5 indicating less than annual performance of the job duty), Subject-matter experts’ mean frequency rating was $M = 1.11$, indicating that the listed job duties were frequently performed.
Discussion

The results of this development study indicate that the legal keyboarding and language arts tests successfully measure the skills that they were designed to assess. Additionally, it appears that the use of these tests is likely appropriate to the selection process of the legal secretary and legal assistant job classifications. However, it is important to note that this development report does not constitute a full content validation study. Such a study would have to account for regional differences, differences in legal specialty, differences in job positions, and differences in specific job work environment. All that can be extrapolated from the present study is that the evaluated legal tests are appropriate to the selection process for the law office in which the testing site was held. The Principles for the Validation and Use of Personnel Selection Procedures (1987) state that full content validation procedures should allow for test administrators to be able to generalize the content validation results to different population samples, something that the current development study does only if it is confirmed through a validation transportability process. Biddle Consulting Group recommends that individuals wishing to uses these tests as a selection device conduct an in-house content validation study. Such a study would ensure that the selection process is fair and applicable to the job environment where the selection process would take place. Coupled with the current development study, which demonstrates basic ability of the instruments to measure the skills that they were designed to measure, administrators of the legal keyboarding and language arts tests will aid many employers in selecting applicants who possess the skill levels needed for acceptable job proficiency.
Development Report for OPAC® System 5.0

Medical Keyboarding and Language Arts Tests

October 1998
Disclaimer

Though the research conducted for this report is thorough and complete, it should in no way be construed as a final validation study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the skill(s) being tested. Because this study was conducted at only one employer, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the skills measured by the tests in this report are essential to the specific job environment in which the in-house validation study was conducted.
Abstract

Medical keyboarding and language arts tests were developed to aid in the selection of properly qualified candidates in the medical assistant and medical secretary job classifications. Three alternative versions of each test were developed. The medical keyboarding test was designed to measure the speed and accuracy of applicants typing medical text. The medical language arts test was designed to measure an applicant’s ability to proofread and spot various grammatical errors in documents that medical assistants and secretaries would typically be expected to analyze and proofread. Three medical industry experts assisted in the development of both tests, and over 20 subject-matter experts participated in the evaluation of the new tests. Eighty-nine percent of medical subject-matter experts who examined the keyboarding test agreed that the test appropriately measured the skill being assessed, and 84% of subject-matter experts who examined the language arts test also agreed that the test appropriately measured the skills being assessed. Medical subject-matter experts were administered all alternative forms of both tests, and their input established alternate form reliability coefficients for each test. Cutoff scores were also derived, based upon the scores of job incumbent subject-matter experts.
Background

The following is a report describing the development process of the OPAC System medical keyboarding and language arts tests. The reason for developing these tests was twofold. First, a product development decision had been made to orient the OPAC System towards the medical industry, as there is a high need for clerical skills in this field and a perceived high demand for skills testing in medical industry. Second, informal feedback from representatives of the medical industry (solicited mainly from tradeshows conventions and telephone interviews) suggested that medical keyboarding and language arts tests might be the most needed tests for the industry, and thus the most likely tests to develop. Additionally, the OPAC System already contained general versions of these tests, so there was both a product history and test format from which to develop the new instruments.

Early Development

Although an informal perusal into the job occupations of medical assistant and secretary revealed that both keyboarding and language arts skills were important to successful performance in these job classifications, more quantifiable evidence needed to be obtained. To that end, 295 health organizations (hospitals, doctor’s offices, etc.) were contacted via facsimile and asked to provide job descriptions for the positions of medical assistant and secretary. Out of the 295 offices contacted, 12 provided complete job descriptions for these positions. All of the received job descriptions indicated that at least some level of minimum competency in the skills of keyboarding and language arts was needed for successful performance in these job classifications. This information provided enough evidence to justify the full development of selection tests measuring keyboarding and language arts skills. Appendix 1 contains all received medical assistant and secretary job descriptions.

Industry Experts

Industry experts were recruited to provide guidance and direction in the test development process. Three industry experts participated in constructing the tests. All industry experts were required to have at least five years of experience in a job classification above the level of medical assistant or secretary (the qualifications of these experts are provided in Appendix 2). It was the duty of the industry experts to first provide materials from which to develop the tests, and then to provide feedback and advice on how to develop the tests. Based on the material provided by industry experts, three alternate versions of each test were developed. Once completed, the tests were shown to industry experts, who then evaluated them as to their content and provided recommended changes. The tests were revised and again presented to industry experts for final approval. Industry experts were compensated for their participation in the test development process.

Test Descriptions

The medical keyboarding test was designed to measure typing speed and accuracy specific to medical documents frequently typed by medical assistants and secretaries. Three alternate versions of the test were constructed. Each version had between 616 and 643 words of text. The text material was selected from actual documents that had been used in a large, Northern California hospital, and the tests were similarly formatted to take into account form, content, and layout of the presented text. All tests were constructed to have roughly the same overall level of difficulty. To distinguish it from regular typing tests, the
The medical keyboarding test contains frequently used medical terminology and other such medical-specific contents (cc, b.i.d., Levothroid, etc). Because of the frequent technical and numeric information contained in the test, it was thought that skill performance differences between the medical keyboarding test and a non-specific keyboarding test might vary, with test takers performing better on a non-specific test (that does not contain the highly technical information found in the medical keyboarding test). In its final format, the medical keyboarding test will be presented to candidates either on a computer screen, or on a hardcopy printout. Appendix 3 contains all three versions of the test.

The medical language arts tests were designed to measure grammar and proofreading skills. As with the medical keyboarding test, three alternate versions of the test were constructed, and these versions were constructed with the intention of being similar in both structure and difficulty level. The test was designed to simulate an actual medical document, such as an insurance claim or doctor’s report. A series of errors were imbedded in the text, the goal for the test taker being to locate and correct these errors. The errors were divided into the classifications of spelling, grammar, punctuation, number usage, possessives, and capitalization. To successfully complete the test, candidates must not only identify the errors (demonstrating proofreading skill), but also have the knowledge to correct the uncovered errors. Each alternate version of the test had between 78 and 82 errors imbedded in the text document, which was between 320 and 348 words long. Error-to-total-word ratios ranged from .23 to .24, which is a similar level found in current OPAC System language arts tests. Appendix 4 contains all three versions of this test.

Testing Site

After construction of the tests was complete, it became necessary to locate a suitable testing site from which to pilot test the new instruments. For the medical keyboarding and language arts tests, a large Health Maintenance Organization located in Roseville, California was selected as the testing site for the new instruments. This test site offered a large pool of subject-matter experts from which to draw, and it also provided subject-matter experts who had some diversity in their particular area of medical expertise. Subject-matter experts from several medical specialties were able to participate in the study.

Method

Participants

Twenty-three subject-matter experts took part in the beta testing of the medical keyboarding and language arts tests (N = 23). All subject-matter experts were either medical assistants or medical secretaries (or of similar classification) and had at least one year of experience working in that job occupation. The overall mean years of job experience for the subject-matter experts was 7.43 (M = 7.43, SD = 6.90). Subject-matter experts spent approximately one hour taking and evaluating all three versions of both tests. Upon completing the test evaluation, subject-matter experts were thanked for their participation and compensated for their time with gift certificates from a local department store.
Materials

Medical Keyboarding and Language Arts Tests.

Final beta versions of the medical keyboarding and language arts tests were administered to subject-matter experts. The tests were contained in a special beta version of OPAC 5.0 skills testing software that had been installed onto a single computer located in the main office of the hospital in which the test site was being held. Candidates were able to open the program by selecting an icon located on the desktop of the computer. Once opened, the computer automatically launched the tests, and candidates completed all three versions of each test.

Validation Survey.

The validation survey was used to evaluate the quality and content validity of each test being examined. The survey was constructed based on a validation report included in OPAC 5.0, and addresses the content validation requirements described in the Uniform Guidelines (1978). Data on each topic was gathered in the survey:

- Whether or not the test measured the skill it was designed to measure
- Whether or not the skill being measured is required at job entry
- The importance of the skill
- The difficulty level of the test
- The subject-matter expert’s score on the test
- The subject-matter expert’s opinion as to what a minimally qualified candidate’s score on the test should be to be considered for employment/promotion

The survey was also designed to capture subject-matter expert demographic information such as name, gender, ethnicity, job title, and years of work experience. All versions of both tests were examined separately, and subject-matter experts completed validation surveys for all versions of each test.

Procedure

An office supervisor at the hospital was placed in charge of the test site. This test proctor arranged individual appointments with each of the subject-matter experts to examine the new medical tests at times that would not interfere with their regular work hours. These individual appointments were staggered out over a two-week period, so as to allow sufficient time for each subject-matter expert to be able to participate. Subject-matter experts were seated at the computer which had the beta version of the OPAC software installed. Once seated, subject-matter experts were given the validation survey, which contained full instructions on how the testing process was to proceed. In order to keep track of their scores on the computer, subject-matter experts entered their social security number when prompted to do so by the computer. The computer then administered each version of both tests to subject-matter experts, who had five minutes to complete each keyboarding test, and 13 minutes to complete each language arts test. The order in which the tests were presented was randomized so as to lessen any carry-over or practice effects. Between each test, the computer was paused, allowing subject-matter experts to answer validation questions about each of the tests in the survey.
After all six tests were completed, the subject-matter experts were asked to attest that they gave each test their best effort, which they did by checking a box on the last page of the survey that indicated as such. Subject-matter experts were thanked for their time and escorted from the test site.

**Results**

In order to establish basic content validity for each test, at least 50% of subject-matter experts must agree that proficiency in the skill which the test measures is essential for successful performance of the job being selected for. Eighty-four percent of subject-matter experts agreed that proficiency in language arts was essential to successful performance in the job of medical assistant or secretary, and 89% agreed that keyboarding skills were necessary for successful performance of the job of medical assistant or secretary.

It is also essential to demonstrate that a skill being tested for is required at the time of job entry, and cannot be learned during a brief orientation. To that end, subject-matter experts were asked whether or not keyboarding and language arts skills were required at time of job entry or if they could be learned while on the job. Seventy-nine percent of subject-matter experts agreed that language arts skills were essential at time of job entry, and 83% percent agreed that keyboarding skills were essential at time of job entry.

**Medical Keyboarding**

Each alternate version of the medical keyboarding test was examined to determine mean scores and difficulty levels for each. Mean scores and standard deviations for the medical keyboarding test versions one, two, and three were highly comparable ($M = 35.78, SD = 11.46, M = 35.09, SD = 13.52, M = 36.72, SD = 12.34$), suggesting that the tests contained similar content and had a similar level of difficulty. The overall mean standard error of measurement was 4.12. In order to determine consistency between the different versions of the test, an alternate form reliability analysis was conducted. The Pearson product-moment correlation coefficient was used to determine the reliability of each version of the test. From this analysis, the following matrix was developed.

**Table 1: Product-moment correlations between each version of the Medical Keyboarding Test.**

<table>
<thead>
<tr>
<th></th>
<th>Medical Keyboarding Version One</th>
<th>Medical Keyboarding Version Two</th>
<th>Medical Keyboarding Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Keyboarding Version One</td>
<td>1.00</td>
<td>.853*</td>
<td>.830*</td>
</tr>
<tr>
<td>Medical Keyboarding Version Two</td>
<td>.853*</td>
<td>1.00</td>
<td>.968*</td>
</tr>
<tr>
<td>Medical Keyboarding Version Three</td>
<td>.830*</td>
<td>.968*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.
Based upon the correlations between each version of the test, an overall mean correlation was determined, $R (22) = .88$, $p < .01$. This is a strong reliability coefficient, and indicates consistency between different versions of the test. Lastly, subject-matter experts were asked to rate the difficulty level of the test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts were asked to rate the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of $M = 2.20$, $SD = .53$, suggesting that the tests are set at an appropriate, if slightly high, difficulty level.

**Medical Language Arts**

As with the medical keyboarding tests, each alternate version of the medical language arts test was examined to determine mean scores and difficulty levels. Mean scores and standard deviations for the medical language arts test versions one, two, and three were consistent ($M = 56.59$, $SD = 13.21$, $M = 59.80$, $SD = 8.29$, $M = 57.67$, $SD = 11.37$), meaning that the tests contained similar content and had a similar level of difficulty. Overall, the mean standard error of measurement was 5.36. As with the medical keyboarding tests, a reliability analysis was conducted. The Pearson product-moment correlation coefficient was again used to determine the reliability of each version of the test. From this analysis, the following matrix was constructed.

**Table 2: Product-moment correlations between each version of the Medical Language Arts Test.**

<table>
<thead>
<tr>
<th></th>
<th>Medical Language Arts Version One</th>
<th>Medical Language Arts Version Two</th>
<th>Medical Language Arts Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Language Arts Version One</td>
<td>1.00</td>
<td>.743*</td>
<td>.745*</td>
</tr>
<tr>
<td>Medical Language Arts Version Two</td>
<td>.743*</td>
<td>1.00</td>
<td>.814*</td>
</tr>
<tr>
<td>Medical Language Arts Version Three</td>
<td>.745*</td>
<td>.814*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.

An overall mean correlation was determined, $R (22) = .77$, $p < .01$. This is an acceptable reliability coefficient, and indicates consistency between different versions of the test. Subject-matter experts were lastly asked to rate the difficulty level of the language arts test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts were asked to rate the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of $M = 1.99$, $SD = 0.50$, indicating that the tests are set at an appropriate level of difficulty.

**Angoff Scores**

To determine the appropriate cutoff score for each test, the modified Angoff method was utilized. The United States Supreme Court (U.S. v. South Carolina) has upheld this method of determining test cutoff scores (Biddle, 1993). Subject-matter experts were asked as to
what they believed the score on each test for a *minimally qualified applicant* should be, which is designed to represent how a minimally qualified job applicant would perform on the test. Subject-matter experts provided these Angoff scores for all versions of each test. Angoff scores were then averaged across alternate versions of each test, yielding a mean Angoff score of 33.92 for the medical keyboarding test, and 56.88 for the medical language arts tests. Based on these Angoff scores, cutoff scores using each test’s standard error of measurement could be derived. The cutoff score for each test was set at one standard error of measurement unit below the test’s mean Angoff. This process led to the following modified Angoff cutoff score for each test.

**Table 3: Summary Statistics and Modified Angoff Cutoff Scores for the Medical Keyboarding and Medical Language Arts Tests.**

<table>
<thead>
<tr>
<th></th>
<th>Medical Keyboarding</th>
<th>Medical Language Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Angoff Score</strong></td>
<td>33.92</td>
<td>56.88</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>12.09</td>
<td>11.12</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>.88</td>
<td>.76</td>
</tr>
<tr>
<td><strong>Mean Standard Error of Measurement</strong></td>
<td>4.12</td>
<td>5.36</td>
</tr>
<tr>
<td><strong>Modified Angoff Cutoff Score</strong></td>
<td>29</td>
<td>51</td>
</tr>
</tbody>
</table>

Appendix 5 contains full summary statistics for each test, as well as raw candidate scores and feedback from each of the selection tests.

**Performance Differentiation**

Lastly, subject-matter experts were polled to determine how strongly they believed that higher levels of mastery in the skill being assessed distinguished candidates with higher levels of performance in a particular job duty from candidates with lower levels of performance in this job duty. Using a Likert-type scale ranging from 1 to 4 (1 indicating little or no performance differentiation, 2 indicating some performance differentiation, 3 indicating significant performance differentiation, and 4, indicating very significant performance differentiation) subject-matter experts were asked to rate how performance differentiating the skills being assessed by the new tests were. Subject-matter experts gave the medical keyboarding test a mean performance differentiation rating of 2.43, and the medical language arts test a mean performance differentiation rating of 2.41, suggesting that higher levels of these skills may be performance differentiating.

**Job Duty/KSA Linkage**

The Uniform Guidelines (1978) require that tested knowledge, skills, and abilities (KSAs) be linked to established job duties. Responses from subject-matter experts almost universally agreed that keyboarding and language arts skills were essential components of major job duties. Subject-matter experts were asked to list the two most important job duties that link to the tested KSAs, and to rank the importance and frequency of each job duty. Job duties such as “typing reports,” “proofreading claims,” and “correct recording of information” were linked to both keyboarding and language arts skills by subject-matter experts. See Appendix 5 for full descriptions. On a Likert-type scale of 1 to 5 (1 being not important, 5 being extremely critical), subject-matter experts rated the overall importance
of listed job duties with a mean rating of $M = 3.38$, indicating that the linked job duties were essential to successful job performance. Subject-matter experts also assigned a frequency rating to the listed job duties. Using a Likert-type scale of 1 to 5 (1 indicating daily to weekly performance of the job duty, 5 indicating less than annual performance of the job duty). Subject-matter experts’ mean frequency rating was $M = 1.20$, indicating that the listed job duties were frequently performed.

**Discussion**

The results of this development study indicate that the medical keyboarding and language arts tests successfully measure the skills that they were designed to assess. Additionally, it appears that the use of these tests is likely appropriate to the selection process of the medical secretary and medical assistant job classifications. However, it is important to note that this development report does not constitute a full content validation study. Such a study would have to account for regional differences, differences in medical specialty, differences in job positions, and differences in specific job work environment. All that can be extrapolated from the present study is that the evaluated medical tests are appropriate to the selection process for the medical office in which the testing site was held. The Principles for the Validation and Use of Personnel Selection Procedures (1987) state that full content validation procedures should allow for test administrators to be able to generalize the content validation results to different population samples, something that the current development study does only if it is confirmed through a validation transportability process. Biddle Consulting Group recommends that individuals wishing to uses these tests as a selection device conduct an in-house content validation study. Such a study would ensure that the selection process is fair and applicable to the job environment where the selection process would take place. Coupled with the current development study, which demonstrates basic ability of the instruments to measure the skills that they were designed to measure, administrators of the medical keyboarding and language arts tests will aid many employers in selecting applicants who possess the skill levels needed for acceptable job proficiency.
Development Report for OPAC® System 5.3
Legal and Medical Transcription Tests

March 1999
Disclaimer

Though the research conducted for this report is thorough and complete, it should in no way be construed as a final validation study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the skill(s) being tested. Because this study was conducted at only one employer, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the skills measured by the tests in this report are essential to the specific job environment in which the in-house validation study was conducted.
Abstract

Legal and medical transcription tests were developed to aid in the selection of properly qualified candidates in the legal secretary and medical transcriptionist job classifications. Three alternative versions of each test were developed. The legal transcription test was designed to measure the ability to accurately transcribe the dictation of a legal document, while the medical transcription test measures medical transcription ability. Two legal industry experts and two medical transcription experts assisted in the development of each test. Thirty-seven subject-matter experts participated in the evaluation of the legal transcription test, while seventeen subject-matter experts evaluated the medical transcription test. Ninety-four percent of legal subject-matter experts who examined the transcription test agreed that the test appropriately measured the skill being assessed, and 100 percent of the medical subject-matter experts who examined the medical transcription test agreed that the test appropriately measured the skills being assessed. All subject-matter experts were administered all alternative forms of both tests, and cutoff scores were derived based upon the performance of job incumbent subject-matter experts.
Background

The following is a report describing the development process of the OPAC System legal and medical transcription tests. The reason for developing these tests was twofold. First, a product development decision had been made to orient the OPAC System towards the legal and medical industries, as there is a high need for clerical skills in these professions and a perceived high demand for skills testing in both the legal and medical industries. Second, informal feedback from representatives of the legal and medical industries (solicited mainly from tradeshows conventions and telephone interviews) suggested that transcription tests might be highly-needs skills tests for both industries, and thus the most likely tests to develop. Additionally, the OPAC System already contained a general version of a transcription test, so there was both a product history and test format from which to develop the new instruments.

Early Development

Although an informal perusal into the job occupations of legal secretary/assistant and medical transcriptionist revealed that transcription skills were important to successful performance in these job classifications, more quantifiable evidence needed to be obtained. To that end, 241 law offices and 295 medical offices throughout the United States were contacted via facsimile and asked to provide job descriptions for the positions of legal secretary/assistant and medical transcriptionist. Out of the 241 law offices contacted, 11 provided complete job descriptions for these positions. Out of the 295 medical offices, 12 provided job descriptions for the position of medical transcription and/or medical secretary. All of the received job descriptions indicated that at least some level of minimum competency in the skill of transcription was needed for successful performance in these job classifications. This information provided enough evidence to justify the full development of selection tests measuring legal and medical transcription ability. Appendix 1 contains all received legal and medical job descriptions.

Industry Experts

Industry experts from both the legal and medical professions were recruited to provide guidance and direction in the test development process. Two industry experts from each profession participated in the construction of the industry-specific transcription tests. All industry experts were required to have at least five years of experience in a job classification at or above the level of legal secretary/assistant or medical transcriptionist (the qualifications of these experts are provided in Appendix 2). It was the duty of the industry experts to first provide materials from which to develop the tests, and then to provide feedback and advice on how to develop the tests. Based on the material provided by industry experts, three alternate versions of each test were developed. Once completed, the tests were shown to industry experts, who then evaluated them as to their content and provided recommended changes. The tests were revised and again presented to industry experts for final approval. Industry experts were compensated for their participation in the test development process.

Test Descriptions

The legal transcription test was designed to assess a candidate’s ability to transcribe dictated legal material that would typically be transcribed by legal secretaries or legal assistants. Three alternate versions of the test were constructed. Each version had a similar
format and structure, and contained between 163 and 194 words imbedded in the text. (The tests could not be constructed to be the exact same word length because of the need to have the each text make structural and grammatical sense.) The text material was selected from actual documents that had been used in a several law offices and the tests were similarly formatted to take into account form, content, and layout of the presented text. All tests were constructed to have roughly the same overall level of difficulty. To distinguish the text from regular transcription material, the legal transcription test contains frequently used legal terminology and other such legal-specific content. In its final format, candidates will transcribe the legal dictation in OPAC using whatever native word processing application is installed on the computer on which they are testing. Appendix 3 contains all three versions of the test.

The medical transcription test was designed to measure skill in medical transcription. Although steps were taken to ensure that the test measures the basic skill of medical transcription, candidates taking this test are also required to have a certain knowledge of medical terminology, as medical transcription often requires transcriptionists to have a rudimentary familiarity with surgical procedures, physiology, pharmacology, etc. As with the legal transcription test, three alternate versions of the test were constructed, and these versions were constructed with the intention of being similar in both structure and difficulty level. The test was designed to simulate actual medical transcriptions, such as a History and Exam Report, or an Operating Report. When taking the test, candidates transcribe from a dictation that is recorded on a cassette tape. As with the legal transcription test, candidates transcribe the dictation into whatever native word processing application is installed on the computer on which they are testing. Appendix 4 contains all three versions of this test.

Testing Site

After construction of the tests was complete, it became necessary to locate a suitable testing site from which to pilot test the new instruments. For the legal transcription test, a large law office located in the San Francisco bay area was selected as the testing site for the new instrument. This test site offered a large pool of subject-matter experts from which to draw, and it also provided subject-matter experts who had some diversity in their particular area of law practice. Subject-matter experts from several fields of law were able to participate in the study.

Medical transcriptionists were gathered from two medical transcription offices located in Sacramento, California. Both of these offices had transcriptionists who specialized in various types of medical transcription, and all of the transcriptionists used had at least several years of experience in the medical transcription profession.

Method

Legal Participants

Thirty-seven subject-matter experts took part in the beta testing of the legal transcription test (N = 37). All subject-matter experts were either legal assistants or legal secretaries (or of similar classification) and had at least one year of experience working in that job occupation. The overall mean years of job experience for the subject-matter experts was 10.86 (M = 10.86, SD = 8.19). Subject-matter experts spent approximately 1/2 hour taking and evaluating all three versions of the test. Upon completing the test evaluation, subject-
matter experts were thanked for their participation and compensated for their time with gift certificates from a local department store.

Medical Participants

Seventeen subject-matter experts took part in the beta testing of the medical transcription test (N = 17). All subject-matter experts were medical transcriptionists (or of similar classification) and had at least one year of experience working in that job occupation. Subject-matter experts spent approximately 1/2 hour taking and evaluating all three versions of the test. Upon completing the test evaluation, subject-matter experts were thanked for their participation and compensated for their time with gift certificates from a local department store.

Materials

Legal and Medical Transcription Tests.

Final beta versions of the legal and medical transcription tests were administered to subject-matter experts. The tests were contained in a special beta-version of the OPAC 5.3 skills testing software that had been installed onto computers in the law office’s training room, offices of the medical transcription service, and at the OPAC office. Candidates were able to open the program by selecting an icon located on the desktop of the computer. Once opened, the computer automatically launched the tests, and candidates completed all three versions of the test.

Dictation Tapes.

Candidates transcribed audio information contained in legal and medical dictation cassettes. The legal dictation cassette contained three audio dictations that legal secretaries would typically be expected to transcribe in a typical law office work environment. The medical dictation cassette contained three dictations that medical transcriptionists frequently transcribe in the course of working in the medical transcription industry. Both tapes were recorded in a professional recording studio to ensure the highest possible quality. The legal dictation tape was recorded with the voice of a professional actor, while the medical transcription dictation was recorded using the voice of a professional medical transcriptionist who was familiar with the verbiage and terminology typically found in medical transcription.

Validation Survey.

The validation survey was used to evaluate the quality and content validity of each test being examined. The survey was constructed based on a validation report included in OPAC 5.0, and addresses the content validation requirements described in the Uniform Guidelines (1978). Data on each topic was gathered in the survey:

- Whether or not the test measured the skill it was designed to measure
- Whether or not the skill being measured is required at job entry
- The importance of the skill
- The difficulty level of the test
- The subject-matter expert’s score on the test
- The subject-matter expert’s opinion as to what a minimally qualified candidate’s score on the test should be to be considered for employment/promotion

Copyright © 1989-2006 Biddle Consulting Group, Inc. 89
The survey was also designed to capture subject-matter expert demographic information such as name, gender, ethnicity, job title, and years of work experience. All versions of both tests were examined separately, and subject-matter experts completed validation surveys for all versions of each test.

Procedure

For the legal transcription test site, a training supervisor at the law office was placed in charge of the test. Subject-matter experts were tested in groups of five or six during their lunch hour. These testing sessions were staggered out over a one-week period, so as to allow sufficient time for each subject-matter expert to be able to participate. For the medical transcription test site, subject-matter experts were tested at both the offices of one of the medical transcription services from which subject-matter experts were recruited, and the offices of OPAC Testing Software. Subject-matter experts were seated at the computer which had the beta version of the OPAC software installed. Once seated, subject-matter experts were given the validation survey, which contained full instructions on how the testing process was to proceed. In order to keep track of their scores on the computer, subject-matter experts entered their social security number when prompted to do so by the computer. The computer then administered all versions of the transcription test to subject-matter experts. Between each test, the computer was paused, allowing subject-matter experts to answer validation questions about each of the tests in the survey.

After all three tests were completed, the subject-matter experts were asked to attest that they gave each test their best effort, which they did by checking a box on the last page of the survey that indicated as such. Subject-matter experts were thanked for their time and escorted from the test site. Subject-matter experts were compensated for their input.

Results

In order to establish basic content validity for each test, at least 50% of subject-matter experts must agree that proficiency in the skill which the test measures is essential for successful performance of the job being selected for. Ninety-four percent of legal subject-matter experts and 100 percent of medical subject-matter experts agreed that proficiency in transcription was essential to successful performance in the job of legal secretary or medical transcriptionist.

It is also essential to demonstrate that a skill being tested for is required at the time of job entry, and cannot be learned during a brief orientation. To that end, subject-matter experts were asked whether or not transcription skills were required at time of job entry or if they could be learned while on the job. Seventy-six percent of legal subject-matter experts and 100 percent of medical subject-matter experts agreed that transcription skills were essential at time of job entry.

Legal Transcription

Each alternate version of the legal transcription test was examined to determine mean scores and difficulty levels for each. Mean scores and standard deviations for the legal transcription test versions one, two, and three were highly comparable (M = 147.63, SD = 26.84, M = 182.22, SD = 28.54, M = 142.47, SD = 22.32), suggesting that the tests contained similar content and had a similar level of difficulty. (The majority of variance around the test means can be attributed to slight differences in the length of each test). The overall mean standard error of measurement was 13.52. In order to determine consistency
between the different versions of the test, an alternate form reliability analysis was conducted. The Pearson product-moment correlation coefficient was used to determine the reliability of each version of the test. From this analysis, the following matrix was developed.

Table 1: Product-moment correlations between each version of the Legal Transcription Test.

<table>
<thead>
<tr>
<th></th>
<th>Legal Transcription Version One</th>
<th>Legal Transcription Version Two</th>
<th>Legal Transcription Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal Transcription</td>
<td>1.00</td>
<td>.935*</td>
<td>.920*</td>
</tr>
<tr>
<td>Version One</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Transcription</td>
<td>.935*</td>
<td>1.00</td>
<td>.983*</td>
</tr>
<tr>
<td>Version Two</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal Transcription</td>
<td>.920*</td>
<td>.983*</td>
<td>1.00</td>
</tr>
<tr>
<td>Version Three</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.

Based upon the correlations between each version of the test, an overall mean correlation was determined, $R (36) = .946$, $p < .01$. This is a strong reliability coefficient, and indicates consistency between different versions of the test. However, in order to produce an internal reliability coefficient for each test (to be used for purposes of constructing a cutoff score) the reliability of the standard OPAC transcription test was applied to the legal transcription test. This is a valid procedure due to the similarity in test design, administration and development between the two tests. The internal reliability coefficient used for the legal transcription test was .7276.

Subject-matter experts were also asked to rate the difficulty level of the test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts rated the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of $M = 1.99$, $SD = 0.55$, indicating that the tests are set at an appropriate level of difficulty.

Medical Transcription

As with the legal transcription test, each alternate version of the medical transcription test was examined to determine mean scores and difficulty levels. Mean scores and standard deviations for the medical transcription test versions one, two, and three were consistent ($M = 69.44$, $SD = 2.45$, $M = 63.88$, $SD = 3.42$, $M = 59.38$, $SD = 7.27$), meaning that the tests contained similar content and had a similar level of difficulty. Overall, the mean standard error of measurement was 2.287. As with the legal transcription test, a reliability analysis was conducted. The Pearson product-moment correlation coefficient was again used to determine the reliability of each version of the test. From this analysis, the following matrix was constructed.
Table 2: Product-moment correlations between each version of the Medical Transcription Test.

<table>
<thead>
<tr>
<th></th>
<th>Medical Transcription Version One</th>
<th>Medical Transcription Version Two</th>
<th>Medical Transcription Version Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Transcription</td>
<td>1.00</td>
<td>.487*</td>
<td>.387</td>
</tr>
<tr>
<td>Version One</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Transcription</td>
<td>.487*</td>
<td>1.00</td>
<td>.571*</td>
</tr>
<tr>
<td>Version Two</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Transcription</td>
<td>.387</td>
<td>.571*</td>
<td>1.00</td>
</tr>
<tr>
<td>Version Three</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.01 level.

An overall mean correlation was determined, \( R (16) = .482, p < .05 \). This is an acceptable reliability coefficient, and indicates consistency between different versions of the test. In order to produce an internal reliability coefficient for each test (to be used for purposes of constructing a cutoff score), the reliability of the standard OPAC transcription test was applied to the medical transcription test. This is a valid procedure due to the similarity in test design, administration and development between the two tests. The internal reliability coefficient used for the medical transcription test was .7276.

Medical subject-matter experts were lastly asked to rate the difficulty level of the language arts test. Using a simple Likert-type scale ranging from 1 to 3 (1 indicating that the test was too easy, 2 indicating that the test had the appropriate level of difficulty, and 3, indicating that the test was too difficult) subject-matter experts rated the overall difficulty of the test. Subject-matter experts rated the tests with a mean difficulty level of \( M = 1.89, SD = .503 \), indicating that the tests are set at an appropriate level of difficulty.

**Angoff Scores**

To determine the appropriate cutoff score for each test, the modified Angoff method was utilized. The United States Supreme Court (U.S. v. South Carolina) has upheld this method of determining test cutoff scores (Biddle, 1993). Subject-matter experts were asked as to what they believed the score on each test for a *minimally qualified applicant* should be, which is designed to represent how a minimally qualified job applicant would perform on the test. Subject-matter experts provided these Angoff scores for all versions of each test. Because the alternate versions of the legal transcription test contained a different number of test items, an Angoff *percentage* had to be derived in order for it to be applicable to all three versions of the test. By dividing the Angoff score of each test version by the number of items in each test version, Angoff percentages were derived. The Angoff percentage for test version one was 64.62 percent, and test version two had a percentage of 62.42 percent. Test version three had an Angoff of 66.69 percent. Because of the close similarity between these percentages, they were collapsed into a general Angoff percentage of 64 percent. The alternate versions of the medical transcription test had the same number of test items, and thus did not require an Angoff percentage.
Angoff scores were averaged across alternate versions of each test, yielding a mean Angoff percentage of 64 percent for the legal transcription test, and an Angoff score of 51 for the medical transcription test. Based on these Angoff scores, cutoff scores using each test’s standard error of measurement could be derived. The cutoff score for each test was set at one standard error of measurement unit below the test’s mean Angoff. This process led to the following modified Angoff cutoff score for each test.

### Table 3: Summary Statistics and Modified Angoff Cutoff Scores for the Legal and Medical Transcription Tests.

<table>
<thead>
<tr>
<th></th>
<th>Legal Transcription</th>
<th>Medical Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Angoff Score</td>
<td>157.817</td>
<td>53.804</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>31.39</td>
<td>4.382</td>
</tr>
<tr>
<td>R</td>
<td>.728</td>
<td>.728</td>
</tr>
<tr>
<td>Mean Standard Error of Measurement</td>
<td>13.52</td>
<td>2.287</td>
</tr>
<tr>
<td>Modified Angoff Cutoff Score/Percentage</td>
<td>64%</td>
<td>51</td>
</tr>
</tbody>
</table>

Appendix 5 contains full summary statistics for each test, as well as raw candidate scores and feedback from each of the selection tests.

**Performance Differentiation**

Lastly, subject-matter experts were polled to determine how strongly they believed that higher levels of mastery in the skill being assessed distinguished candidates with higher levels of performance in a particular job duty from candidates with lower levels of performance in the job duty. Using a Likert-type scale ranging from 1 to 4 (1 indicating little or no performance differentiation, 2 indicating some performance differentiation, 3 indicating significant performance differentiation, and 4, indicating very significant performance differentiation) subject-matter experts were asked to rate how performance-differentiating the skills being assessed by the new tests were. Legal subject-matter experts gave the legal transcription test a mean performance differentiation rating of 2.54, and medical subject-matter experts gave the medical transcription test a mean performance differentiation rating of 3.46, suggesting that higher levels of these skills may be performance differentiating.

**Job Duty/KSA Linkage**

The Uniform Guidelines (1978) require that tested knowledge, skills, and abilities (KSAs) be linked to established job duties. Responses from subject-matter experts almost universally agreed that transcription skills were essential components of the major job duties of legal and medical transcription. Subject-matter experts were asked to list the two most important job duties that link to the tested KSAs, and to rank the importance and frequency of each job duty. Job duties such as “processing of documents,” “transcription,” and “drafting correspondence” were linked to both transcription tests by subject-matter experts. See Appendix 5 for full descriptions. On a Likert-type scale of 1 to 5 (1 being not important, 5 being extremely critical), subject-matter experts rated the overall importance of listed job duties for legal transcription with a mean rating of $M = 3.44$, indicating that the linked job duties were essential to successful job performance as a legal secretary. Medical subject-matter experts also assigned an importance rating to the listed job duties. Using a Likert-type scale of 1 to 5 (1 being not important, 5 being extremely critical), subject-matter
experts’ mean importance rating was $M = 4.56$, indicating that the listed job duties were critical for successful performance as a medical transcriptionist.

**Discussion**

The results of this development study indicate that the legal and medical transcription tests successfully measure the skills that they were designed to assess. Additionally, it appears that the use of these tests is likely appropriate to the selection process of the legal secretary and medical transcriptionist job classifications. However, it is important to note that this development report does not constitute a full content validation study. Such a study would have to account for regional differences, differences in job specialty, differences in job positions, and differences in specific job work environment. All that can be extrapolated from the present study is that the evaluated tests are appropriate to the selection process for the office in which the testing site was held. The Principles for the Validation and Use of Personnel Selection Procedures (1987) state that full content validation procedures should allow for test administrators to be able to generalize the content validation results to different population samples, something that the current development study does only if it is confirmed through a validation transportability process. Biddle Consulting Group recommends that individuals wishing to use these tests as a selection device conduct an in-house content validation study. Such a study would ensure that the selection process is fair and applicable to the job environment where the selection process would take place. Coupled with the current development study, which demonstrates basic ability of the instruments to measure the skills that they were designed to measure, administrators of the legal and medical transcription tests will aid many employers in selecting applicants who possess the skill levels needed for acceptable job proficiency.
References


Contreras v. City of Los Angeles, 656 F.2d 1267 (9th Cir. 1981).


Waisome v. Port Authority of New York, 948 F.2d 1370 (2nd Cir. 1991).
Development and Research Report
for OPAC® System 7.0
PowerPoint Test

April 2002
OPAC® PowerPoint Test

In March of 2002, an internal consistency reliability study of a beta version of the OPAC® PowerPoint Test was conducted. The reliability was found to be .72. The United States Department of Labor's general guidelines for interpreting reliability coefficients indicate that this level of reliability is interpreted as being “adequate.” Also, according to these guidelines the OPAC Intermediate PowerPoint Test has sufficient reliability to be used as a selection device if it is also valid for the position being tested.

About the Test

The OPAC® PowerPoint Test measures a person’s ability to correctly use important features found in the Microsoft® PowerPoint program. The areas measured during the test are based upon input received from an Industrial and Organizational Psychologist, who has an extensive background in training others how to use Microsoft Office products at the basic, intermediate, and advanced levels. It is also inspired by objectives set forth by Microsoft for Microsoft Office Specialist Skill Standards: PowerPoint 2000. Test takers must be familiar with terminology specifically associated with the Microsoft® PowerPoint program to be able to accurately respond to the test items.

The Industrial and Organizational Psychologist who assisted in the development of this test is an Assistant Professor of Administration at a major university. This psychologist is certified by Microsoft as a Microsoft Office User Specialist Authorized Instructor and has worked as a subject-matter expert to NIVO International (the manufacturer of the Microsoft Certification exams). In addition, she is also certified by Microsoft at the Master Level in Microsoft Office products.

According to the Microsoft web site, the Microsoft Office Specialist PowerPoint 2000 exam was created and validated by industry experts, and Microsoft’s exam development process is accredited by the Buros Institute for Assessment Consultation and Outreach. A complete listing of the PowerPoint exam-skill standards published by Microsoft can be found at http://www.microsoft.com/learning/mcp/officespecialist/objectives/excel2000.asp.

Test Reliability Study Participants

Twenty-one office workers who indicated that they possessed at least a basic knowledge and understanding of the PowerPoint program took a beta version of the OPAC PowerPoint test in March of 2002 at Biddle Consulting Group, Inc.’s corporate offices in Rancho Cordova, CA.

Descriptive Statistics, Including Reliability, for OPAC Intermediate PowerPoint Test [Section 15C(5)]

The following are the descriptive statistics for the OPAC PowerPoint Test (beta version) including measures of central tendency (i.e., means), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines. Alpha (internal consistency; Cronbach, 1951) reliability analysis was conducted for this test using SPSS 13.0. The mean and standard deviation of these three test modules are provided in percentage score.
<table>
<thead>
<tr>
<th>Microsoft PowerPoint Test</th>
<th>Sample Size</th>
<th># of Test Items</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Internal Consistency (alpha) reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>25</td>
<td>73.14% 12.78% .7199</td>
</tr>
</tbody>
</table>

The U. S. Department of Labor indicates that reliability coefficients of 0.70 – 79 is to be interpreted as being “adequate.”

### Validity

The Validation Wizard, which is included with the OPAC software, is designed to help users who are not testing experts address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected groups. However, even for tests without adverse impact, it make good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. If you find that this test adversely impacts a protected group of test takers, it is important to ensure it meets validity standards. Call Biddle Consulting Group, Inc. toll free at 800-999-0438 for consulting advice about Job Analyses, Validation, and Test Fairness.

### Accuracy and Completeness [Section 15C(9)]

Biddle Consulting Group, Inc consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees (or contractors) and then independently checked for accuracy by trained BCG employees. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test in this report are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Test administrators can conduct in-house validation studies using a service provided by Biddle Consulting Group, Inc.

### Contact Person [Section 15C(8)]

To receive further information about this study, contact:

Biddle Consulting Group, Inc.
Attention: James E. Kuthy
270 Blue Ravine Road, Suite 270

Copyright © 1989-2006 Biddle Consulting Group, Inc.
OPAC® and Office Proficiency and Assessment Certification® are registered trademarks of Biddle Consulting Group, Inc.

Microsoft® is a registered trademark or trademark of the Microsoft Corporation in the United States and/or other countries. The OPAC® Intermediate PowerPoint Test is an independent publication of Biddle Consulting Group, Inc. and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation.

References


Development and Research Report
for OPAC® 7.0
Intermediate Excel Test

April 2002
OPAC® Intermediate Excel Test

In March of 2002, an internal consistency reliability study of a beta version of the OPAC® Intermediate Excel Test was conducted. The reliability was found to be .75. The United States Department of Labor’s general guidelines for interpreting reliability coefficients indicate that this level of reliability is interpreted as being “adequate.” Also, according to these guidelines the OPAC Intermediate Excel Test has sufficient reliability to be used as a selection device if it is also valid for the position being tested.

About the Test

The OPAC® Intermediate Excel Test measures a person’s ability to correctly use important intermediate-level features found in the Microsoft® Excel program. The areas measured during the test are based upon input received from an Industrial and Organizational Psychologist, who has an extensive background in training others how to use Microsoft Office products at the basic, intermediate, and advanced levels. It is also inspired by objectives set forth by Microsoft for Microsoft Office Specialist Skill Standards: Excel 2000 Expert. Test takers must be familiar with terminology specifically associated with the Microsoft® Excel program to be able to accurately respond to the test items.

The Industrial and Organizational Psychologist who assisted in the development of this test is an Assistant Professor of Administration at a major university. This psychologist is certified by Microsoft as a Microsoft Office User Specialist Authorized Instructor and has worked as a subject-matter expert to NIVO International (the manufacturer of the Microsoft Certification exams). In addition, she is also certified by Microsoft at the Master Level in Microsoft Office products and at the Expert Level in Microsoft Excel.

According to the Microsoft web site, the Microsoft Office Specialist Excel 2000 Expert exam was created and validated by industry experts, and Microsoft’s exam development process is accredited by the Buros Institute for Assessment Consultation and Outreach. A complete listing of the Excel exam-skill standards published by Microsoft can be found at http://www.microsoft.com/learning/mcp/officespecialist/objectives/excel2000.asp.

Test Reliability Study Participants

Twenty-seven office workers who indicated that they possessed at least a basic knowledge and understanding of the Excel program took a beta version of the OPAC Intermediate Excel test in March of 2002 at Biddle Consulting Group, Inc.’s corporate offices in Rancho Cordova, CA.

Descriptive Statistics, Including Reliability, for OPAC Intermediate Excel Test [Section 15C(5)]

The following are the descriptive statistics for the OPAC Intermediate Excel Test (beta version) including measures of central tendency (i.e., means), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines. Alpha (internal consistency; Cronbach, 1951) reliability analysis was conducted for this test using SPSS 13.0. The mean and standard deviation of these three test modules are provided in raw score.
The U. S. Department of Labor indicates that reliability coefficients of 0.70 – 79 is to be interpreted as being “adequate.”

**Validity**

The **Validation Wizard**, which is included with the OPAC software, is designed to help users who are not testing experts address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected groups. However, even for tests without adverse impact, it make good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. If you find that this test adversely impacts a protected group of test takers, it is important to ensure it meets validity standards. Call Biddle Consulting Group, Inc. toll free at 800-999-0438 for consulting advice about Job Analyses, Validation, and Test Fairness.

**Accuracy and Completeness [Section 15C(9)]**

Biddle Consulting Group, Inc consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees (or contractors) and then independently checked for accuracy by trained BCG employees. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test in this report are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Test administrators can conduct in-house validation studies using a service provided by Biddle Consulting Group, Inc.

**Contact Person [Section 15C(8)]**

To receive further information about this study, contact:

Biddle Consulting Group, Inc.
Attention: James E. Kuthy
270 Blue Ravine Road, Suite 270

Copyright © 1989-2006 Biddle Consulting Group, Inc.
OPAC® and Office Proficiency and Assessment Certification® are registered trademarks of Biddle Consulting Group, Inc.

Microsoft® is a registered trademark or trademark of the Microsoft Corporation in the United States and/or other countries. The OPAC® Intermediate Excel Test is an independent publication of Biddle Consulting Group, Inc. and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation.

References


Development and Research Report
for OPAC® System7.5
Intermediate Word Test

April, 2003
OPAC® Intermediate Word Test

In March of 2003, an internal consistency reliability study of a beta version of the OPAC® Intermediate Word Test was conducted. The reliability was found to be .82. The United States Department of Labor’s general guidelines for interpreting reliability coefficients indicate that this level of reliability is interpreted as being “Good.” Also, according to these guidelines the OPAC Intermediate Word Test has sufficient reliability to be used as a selection device if it is also valid for the position being tested.

About the Test

The OPAC® Intermediate Word Test measures a person’s ability to correctly use important intermediate-level features found in the Microsoft® Word program. The areas measured during the test are based upon input received from an Industrial and Organizational Psychologist, who has an extensive background in training others how to use Microsoft Office products at the basic, intermediate, and advanced levels. It is also inspired by objectives set forth by Microsoft for Microsoft Office Specialist Skill Standards: Word 2000 Expert. Test takers must be familiar with terminology specifically associated with the Microsoft® Word program to be able to accurately respond to the test items.

The Industrial and Organizational Psychologist who assisted in the development of this test is an Assistant Professor of Administration at a major university. This psychologist is certified by Microsoft as a Microsoft Office User Specialist Authorized Instructor and has worked as a subject-matter expert to NIVO International (the manufacturer of the Microsoft Certification exams). In addition, she is also certified by Microsoft at the Master Level in Microsoft Office products and at the Expert Level in Microsoft Word.

According to the Microsoft web site, the Microsoft Office Specialist Word 2000 Expert exam was created and validated by industry experts, and Microsoft’s exam development process is accredited by the Buros Institute for Assessment Consultation and Outreach. A complete listing of the Word exam-skill standards published by Microsoft can be found at http://www.microsoft.com/learning/mcp/officespecialist/objectives/excel2000.asp.

Test Reliability Study Participants

Forty-three office workers who indicated that they possessed at least a basic knowledge and understanding of the Word program took a beta version of the OPAC Intermediate Word test in March of 2003 at Biddle Consulting Group, Inc.’s corporate offices in Rancho Cordova, CA.

Descriptive Statistics, Including Reliability, for OPAC Intermediate Word Test [Section 15C(5)]

The following are the descriptive statistics for the OPAC Intermediate Word Test (beta version) including measures of central tendency (i.e., means), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines. Alpha (internal consistency; Cronbach, 1951) reliability analysis was conducted for this test using SPSS 13.0. The mean and standard deviation of these three test modules are provided in percentage score.
The U. S. Department of Labor indicates that reliability coefficients of 0.80 – 89 is to be interpreted as being “good.”

**Validity**

The **Validation Wizard**, which is included with the OPAC software, is designed to help users who are not testing experts address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected groups. However, even for tests without adverse impact, it make good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. If you find that this test adversely impacts a protected group of test takers, it is important to ensure it meets validity standards. Call Biddle Consulting Group, Inc. toll free at 800-999-0438 for consulting advice about Job Analyses, Validation, and Test Fairness.

**Accuracy and Completeness [Section 15C(9)]**

Biddle Consulting Group, Inc consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees (or contractors) and then independently checked for accuracy by trained BCG employees. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test in this report are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Test administrators can conduct in-house validation studies using a service provided by Biddle Consulting Group, Inc.

**Contact Person [Section 15C(8)]**

To receive further information about this study, contact:

Biddle Consulting Group, Inc.
Attention: James E. Kuthy
270 Blue Ravine Road, Suite 270
Folsom, CA 95630

Copyright © 1989-2006 Biddle Consulting Group, Inc.
106
OPAC® and Office Proficiency and Assessment Certification® are registered trademarks of Biddle Consulting Group, Inc.

Microsoft® is a registered trademark or trademark of the Microsoft Corporation in the United States and/or other countries. The OPAC® Intermediate Word Test is an independent publication of Biddle Consulting Group, Inc. and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation.

References


Development and Research Report
For OPAC® System 8.0
Basic Word Test

May 2005
OPAC® Basic Word Test

In April of 2005, an internal consistency reliability study of a beta version of the OPAC® Basic Word Test was conducted. The reliability was found to be .916. The United States Department of Labor's general guidelines for interpreting reliability coefficients indicate that this level of reliability is interpreted as being “excellent.” Also, according to these guidelines the OPAC Basic Word Test has sufficient reliability to be used as a selection device if it is also valid for the position being tested. Based on findings from the present study the final version of the test was improved and is likely to have even a higher validity coefficient than was previously found.

About the Test

The OPAC® Basic Word Test measures a person’s ability to correctly use important basic features found in the Microsoft® Word program at a basic level. The areas measured during the test are based upon input received from an Industrial and Organizational Psychologist, who has an extensive background in training others how to use Microsoft Office products at both the basic and advanced levels. It is also inspired by objectives set forth by Microsoft for their Word 2000 “Office Specialist” examination. Test takers must be familiar with terminology specifically associated with the Microsoft® Word program to be able to accurately respond to the test items.

The Industrial and Organizational Psychologist who assisted in the development of this test is an Assistant Professor of Administration at a major university. This psychologist is certified by Microsoft as a Microsoft Office User Specialist Authorized Instructor and has worked as a subject-matter expert to NIVO International (the manufacturer of the Microsoft Certification exams). In addition, she is also certified by Microsoft at the Master Level in Microsoft Office products and at the Expert Level in Microsoft Word.

According to the Microsoft web site, the Microsoft Office Specialist Word 2000 exam was created and validated by industry experts, and Microsoft’s exam development process is accredited by the Buros Institute for Assessment Consultation and Outreach. A complete listing of the Word exam-skill standards published by Microsoft can be found at http://www.microsoft.com/learning/mcp/officespecialist/objectives/word2000.asp.

Test Reliability Study Demographics

Thirty-eight people who indicated that they possessed at least a basic knowledge and understanding of the Word program took a beta version of the OPAC Basic Word test in April 2005. Twenty-five of those were administered the computerized test at an Adult Learning Center run by a local school district in Sacramento, California. The remaining 12 were administered the same test at Biddle Consulting Group, Inc.’s corporate offices in Rancho Cordova, CA.

The gender of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>33</td>
</tr>
</tbody>
</table>

Copyright © 1989-2006 Biddle Consulting Group, Inc.
The race/ethnicity of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>White</th>
<th>Black/African American</th>
<th>Hispanic/Latino</th>
<th>Asian/Pacific Islander</th>
<th>Native American/Alaska Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

The age of the persons that participated in the current study was:

<table>
<thead>
<tr>
<th>Less than 20 years of age</th>
<th>20-29 years of age</th>
<th>30-39 years of age</th>
<th>40-49 years of age</th>
<th>50 and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

The education level of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>Less than High School Graduate</th>
<th>High School Graduate</th>
<th>GED Certificate</th>
<th>Less than 2-Years College</th>
<th>2-Year College</th>
<th>4-Year College</th>
<th>Graduate Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>11</td>
<td>2</td>
</tr>
</tbody>
</table>

### Descriptive Statistics, Including Reliability, for OPAC Basic Word Test [Section 15C(5)]

The following are the descriptive statistics for the OPAC Basic Word Test (beta version) including measures of central tendency (i.e., means), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines. Alpha (internal consistency; Cronbach, 1951) reliability analysis was conducted for this test using SPSS 13.0. The mean and standard deviation of these three test modules are provided in raw score.

<table>
<thead>
<tr>
<th>Basic Microsoft Word Test</th>
<th>Sample Size</th>
<th># of Test Items</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Internal Consistency (alpha) reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>38</td>
<td>25</td>
<td>12.89</td>
<td>6.38</td>
<td>.916</td>
</tr>
<tr>
<td>Office Administration</td>
<td>13</td>
<td>25</td>
<td>17.92</td>
<td>4.07</td>
<td>- - -</td>
</tr>
<tr>
<td>Learning Center Administration</td>
<td>25</td>
<td>25</td>
<td>10.28</td>
<td>5.81</td>
<td>- - -</td>
</tr>
</tbody>
</table>

Note: The “Learning Center” participants appear to have had extremely limited ability to properly use the Microsoft Word program. For example, only 36% from this venue were able to correctly insert a page break during the test, while 100% of the participants from the “office administration” group correctly performed this function.

A relatively strong, significant relationship was found between the self-reported level of keyboarding speed and test scores ($r = .762$). In other words, those who reported being able to type more quickly typically scored higher on the test.

After the reliability study was conducted modifications were made to the wording of some of the test items based upon the feedback received from those who took the beta version of the test. Follow up testing of several of the test takers from the “office
administration” group revealed that these modifications typically raised their score by two to three points. However, their relatively higher initial scores might have limited the possible improvement to their scores. Based on this finding it is anticipated that both the mean and the reliability of the final version of the test will be somewhat higher than was found during this study and that the standard deviation will likely be reduced.

**Validity**

The **Validation Wizard**, which is included with the OPAC software, is designed to help users who are not testing experts address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected groups. However, even for tests without adverse impact, it make good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. If you find that this test adversely impacts a protected group of test takers, it is important to ensure it meets validity standards. Call Biddle Consulting Group, Inc. toll free at 800-999-0438 for consulting advice about Job Analyses, Validation, and Test Fairness.

**Accuracy and Completeness [Section 15C(9)]**

Biddle Consulting Group, Inc consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees (or contractors) and then independently checked for accuracy by trained BCG employees. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test in this report are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Test administrators can conduct in-house validation studies using a service provided by Biddle Consulting Group, Inc.

**Contact Person [Section 15C(8)]**

To receive further information about this study, contact:

Biddle Consulting Group, Inc.
Attention: James E. Kuthy
193 Blue Ravine Road, Suite 270
Folsom, CA 95630
Voice (916) 294-4250 · Fax (916) 294-4255
OPAC® and Office Proficiency and Assessment Certification® are registered trademarks of Biddle Consulting Group, Inc.

Microsoft® is a registered trademark or trademark of the Microsoft Corporation in the United States and/or other countries. The OPAC® Basic Word Test is an independent publication of Biddle Consulting Group, Inc. and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation.

References


Development and Research Report
for OPAC® System 8.0
Basic Excel Test

May, 2005
OPAC® Basic Excel Test

In April of 2005, an internal consistency reliability study of a beta version of the OPAC® Basic Excel Test was conducted. The reliability was found to be .950. The United States Department of Labor’s general guidelines for interpreting reliability coefficients indicate that this level of reliability is interpreted as being “Excellent.” Also, according to these guidelines the OPAC Basic Excel Test has sufficient reliability to be used as a selection device if it is also valid for the position being tested. Based on findings from the present study the final version of the test was improved and is likely to have even a higher validity coefficient than was previously found.

About the Test

The OPAC® Basic Excel Test measures a person’s ability to correctly use important basic features found in the Microsoft® Excel program at a basic level. The areas measured during the test are based upon input received from an Industrial and Organizational Psychologist, who has an extensive background in training others how to use Microsoft Office products at both the basic and advanced levels. It is also inspired by objectives set forth by Microsoft for their Word Excel “Office Specialist” examination. Test takers must be familiar with terminology specifically associated with the Microsoft® Excel program to be able to accurately respond to the test items.

The Industrial and Organizational Psychologist who assisted in the development of this test is an Assistant Professor of Administration at a major university. This psychologist is certified by Microsoft as a Microsoft Office User Specialist Authorized Instructor and has worked as a subject-matter expert to NIVO International (the manufacturer of the Microsoft Certification exams). In addition, she is also certified by Microsoft at the Master Level in Microsoft Office products and at the Expert Level in Microsoft Excel.

According to the Microsoft web site, the Microsoft Office Specialist Word 2000 exam was created and validated by industry experts, and Microsoft’s exam development process is accredited by the Buros Institute for Assessment Consultation and Outreach. A complete listing of the Excel exam-skill standards published by Microsoft can be found at http://www.microsoft.com/learning/mcp/officespecialist/objectives/excel2000.asp.

Test Reliability Study Demographics

Thirty-five people who indicated that they possessed at least a basic knowledge and understanding of the Word program took a beta version of the OPAC Basic Excel test in April 2005. Twenty-two of those were administered the computerized test at an Adult Learning Center run by a local school district in Sacramento, California. The remaining 13 were administered the same test at Biddle Consulting Group, Inc.’s corporate offices in Rancho Cordova, CA.

The gender of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>32</td>
</tr>
</tbody>
</table>
The race/ethnicity of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>White</th>
<th>Black/African American</th>
<th>Hispanic/Latino</th>
<th>Asian/Pacific Islander</th>
<th>Native American/Alaska Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

The age of the persons that participated in the current study was:

<table>
<thead>
<tr>
<th>Less than 20 years of age</th>
<th>20-29 years of age</th>
<th>30-39 years of age</th>
<th>40-49 years of age</th>
<th>50 and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

The education level of the persons who participated in the current study was:

<table>
<thead>
<tr>
<th>Less than High School Graduate</th>
<th>High School Graduate</th>
<th>GED Certificate</th>
<th>Less than 2-Years College</th>
<th>2-Year College</th>
<th>4-Year College</th>
<th>Graduated Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

**Descriptive Statistics, Including Reliability, for OPAC Basic Excel Test [Section 15C(5)]**

The following are the descriptive statistics for the OPAC Basic Excel Test (beta version) including measures of central tendency (i.e., means), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines. Alpha (internal consistency; Cronbach, 1951) reliability analysis was conducted for this test using SPSS 13.0. The mean and standard deviation of these three test modules are provided in raw score.

<table>
<thead>
<tr>
<th>Basic Microsoft Word Test</th>
<th>Sample Size</th>
<th># of Test Items</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Internal Consistency (alpha ) reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>35</td>
<td>25</td>
<td>12.371</td>
<td>7.674</td>
<td>.950</td>
</tr>
<tr>
<td>Office Administration</td>
<td>13</td>
<td>25</td>
<td>19.308</td>
<td>2.463</td>
<td>- - -</td>
</tr>
<tr>
<td>Learning Center Administration</td>
<td>22</td>
<td>25</td>
<td>8.273</td>
<td>6.670</td>
<td>- - -</td>
</tr>
</tbody>
</table>

Note: The “Learning Center” participants appear to have had limited ability to properly use the Microsoft Excel program. For example, only 50% from this venue were able to correctly re-name a worksheet, while 100% of the participants from the “office administration” group correctly performed this function.

The U. S. Department of Labor indicates that reliability coefficients of 0.90 and higher are interpreted as being “excellent.”

**Validity**

The Validation Wizard, which is included with the OPAC software, is designed to help users who are not testing experts address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected
groups. However, even for tests without adverse impact, it make good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. If you find that this test adversely impacts a protected group of test takers, it is important to ensure it meets validity standards. Call Biddle Consulting Group, Inc. toll free at 800-999-0438 for consulting advice about Job Analyses, Validation, and Test Fairness.

Accuracy and Completeness [Section 15C(9)]

Biddle Consulting Group, Inc consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees (or contractors) and then independently checked for accuracy by trained BCG employees. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been pilot tested, and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test in this report are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Test administrators can conduct in-house validation studies using a service provided by Biddle Consulting Group, Inc.

Contact Person [Section 15C(8)]

To receive further information about this study, contact:

Biddle Consulting Group, Inc.
Attention: James E. Kuthy
193 Blue Ravine Road, Suite 270
Folsom, CA 95630
Voice (916) 294-4250 · Fax (916) 294-4255

OPAC® and Office Proficiency and Assessment Certification® are registered trademarks of Biddle Consulting Group, Inc.

Microsoft® is a registered trademark or trademark of the Microsoft Corporation in the United States and/or other countries. The OPAC® Basic Excel Test is an independent publication of Biddle Consulting Group, Inc. and is not affiliated with, nor has it been authorized, sponsored, or otherwise approved by Microsoft Corporation.
References


Development and Research Report
for OPAC System 8.5
Contemporary Keyboarding Test
(Keyboarding 2)

October 2005
OPAC® Contemporary Keyboarding Test
(Keyboarding 2)

The OPAC Contemporary Keyboarding Test is an extremely realistic, cutting-edge measure of a person’s ability to quickly and accurately copy and enter often-difficult text using a keyboard. According to the U. S. Department of Labor’s Guidelines, it has excellent reliability. It contains many unique features that differentiate it from other keyboarding tests currently being offered. For example, the letter frequency on each of the test versions is, on average, within 98% of the frequency as indicated in the letters occurring in the words listed in the main entries of the Concise Oxford Dictionary (9th edition, 1995). In addition, the content of each of the three test versions also contains:

- 35 numerals
- 10 symbols that are found above the numbers on a keyboard
- Instances of consecutive all-caps
- An assortment of punctuation marks
- Words with repeated letters
- Mixtures of long and short sentences
- Mixtures of long and short words
- Title case words forcing typists to capitalize the first letter of several consecutive words
- Words that should be unfamiliar to the typist, thus a better measure of letter processing speed as opposed to measuring spelling ability
- Grammatically correct phrases
- Realistic to a modern business setting
- Cutting edge
  - Website address(es)
  - Email address(es)
  - Package tracking alphanumeric code(s)
  - Business appropriate terms such as “per diem”

Test Description

The Contemporary Keyboarding Test is a timed test of a person’s ability to read and enter information into a computer using a keyboard or other input device. Each version of this test has the test-taker read and enter information for five (5) minutes, following a one (1) minute warm-up practice test of similar difficulty.

Test scoring is computed using the following calculations:
- Gross WPM = (Gross Keystrokes / 5 Keystrokes per Word) / # of Minutes in Test
- Net WPM = (Net Keystrokes / 5 Keystrokes per Word) / # of Minutes in Test
- Accuracy Rate = Net Keystrokes / Gross Keystrokes

When calculating Net Keystrokes, the OPAC System subtracts the error keystrokes from the gross keystrokes. Each incorrect word (misspelled word, missing word, or added word) counts five error keystrokes, and five keystrokes constitute one word.

---

16 Test-takers’ reported keyboarding speed is likely to be slower when using this test than when measured using more traditional-style keyboarding tests since this test has been rated as being more difficult than traditional-style keyboarding tests. The average Flesch-Kincaid readability grade level of this test is 11.8.
Test Administration for Reliability Study

The Contemporary Keyboarding Test (Keyboarding 2) was administered twice to each of the participants in the current study during August and September 2005. Seventeen of the tests were administered at Biddle Consulting Group, Inc.’s corporate offices in Folsom, California. The test was also administered to ten participants at an adult learning center in Milan, Ohio, and to another ten participants at an Adult Learning Center in San Jose, California.

Descriptive Statistics, Including Reliability, for the OPAC Contemporary Keyboarding Test (Keyboarding 2) [Uniform Guidelines Section 15C(5)17]

The following are the descriptive statistics for the OPAC Contemporary Keyboarding Test (Keyboarding 2) including measures of central tendency (i.e., means/averages), dispersion (i.e., standard deviations), and estimates of reliability as specified by Section 14[C](5) of the federal Guidelines, along with the standard error of measurement. The mean and standard deviation of these three test versions are provided in a Word-Per-Minute (WPM) metric.

<table>
<thead>
<tr>
<th>OPAC Keyboarding 2 Test</th>
<th>Version 1</th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flesch-Kincaid Reading Grade Level</td>
<td>12.0</td>
<td>12.0</td>
<td>11.6</td>
</tr>
<tr>
<td>Average Net WPM18</td>
<td>46.900</td>
<td>46.450</td>
<td>45.233</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>14.797</td>
<td>15.156</td>
<td>13.069</td>
</tr>
<tr>
<td>Reliability</td>
<td>0.976</td>
<td>0.958</td>
<td>0.985</td>
</tr>
<tr>
<td>Standard Error of Measurement19</td>
<td>2.295</td>
<td>3.093</td>
<td>1.597</td>
</tr>
<tr>
<td>Sample Size</td>
<td>35</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

The U. S. Department of Labor indicates that reliability coefficients of 0.90 and higher are interpreted as being “excellent.” Finally, the three versions of the test appear to be extremely parallel in difficulty since there was less than a two Word-Per-Minute average net-score difference between the test versions (i.e., 46.900, 46.450, and 45.233 Words-Per-Minute).

The following chart shows the inter-correlation between the three versions of the Contemporary Keyboarding Test. As can be seen here, all of these tests are strongly correlated with one another (i.e., significant at the $p < .01$ level).

---

17 The section numbers listed in this document refer to sections of the federal Uniform Guidelines on Employee Selection Procedure.
18 The average net WPM score and standard deviation were calculated using scores from the first of the two test administrations.
19 The SEM was calculated using the test-retest reliability of Version 1 and the standard deviation from Version 1 of the first of the two test administrations.

Copyright © 1989-2006 Biddle Consulting Group, Inc.

120
Validity

The Validation Wizard, which is included with the OPAC software, is designed for conducting a very basic content-validity analysis of an OPAC test. Using this feature a test administration can determine if a test measures specific job-related knowledge, skills, or abilities for particular job classifications. It is designed to help users who are not testing experts to address minimum standards of job relatedness for tests which are anticipated or known to produce little, if any, adverse impact on protected groups. However, even for tests without adverse impact, it makes good business sense to establish their job relatedness in order to be fair to candidates and to obtain employees who have adequate levels of knowledge, skills, and abilities actually needed on the job. Biddle Consulting Group, Inc. can help those users who wish to conduct more in-depth validity or reliability analyses of their pre-employment testing.

Accuracy and Completeness [Uniform Guidelines Section 15C(9)]

Biddle Consulting Group, Inc. consultants and staff conducted the study from which the reliability findings reported in this document were collected. The data collected was entered by administrative staff employees and then independently checked for accuracy. Analyses were also independently double-checked and verified. We invite any comments you might have about this report.

Potential Limitations of this Study

Though the research conducted for this report is accurate and complete, it should in no way be construed as a final study. Rather, it is a good faith effort on the part of Biddle Consulting Group, Inc., to demonstrate that the tests described in this report have been “pilot tested,” and that they do provide a meaningful measurement of the abilities and skill(s) being tested. Because this study was conducted as part of an on-going test development process, and included participants from positions that may be dissimilar to those in other organizations, its results and applications may or may not be relevant in other geographical areas, employers, specific areas of practice, or job positions. Biddle

**. Correlation is significant at the 0.01 level (2-tailed).
Consulting Group recommends conducting an in-house validation study of all tests before using them as a selection device, as such a study would help establish that the abilities and skills measured by the test are essential to the specific job environment in which the in-house validation study was conducted. Conducting an in-house study will also evaluate whether the use of the scores (i.e., pass/fail, banding, or ranking) are appropriate for the position(s) in your organization. Biddle Consulting Group, Inc. can assist organizations in conducting job analysis and test validation studies.

**Contact Person [Uniform Guidelines Section 15C(8)]**

To receive further information about this study, contact:
Biddle Consulting Group, Inc.
Attention: James E. Kuthy
193 Blue Ravine Road, Suite 270
Folsom, CA 95630
Voice (916) 294-4250 · Fax (916) 294-4255

**References**
