Job matching assessment: Inter-rater reliability of an instrument assessing employment characteristics of young adults with intellectual disabilities

Robert L. Morgan
Department of Special Education and Rehabilitation, Utah State University, 2865 Old Main, Logan, UT 84322-2865, USA
Tel.: +1 84322 2865; Fax: +1 435 797 3251; E-mail: bob.morgan@usu.edu

Revised/Accepted: November 2010

Abstract. Job matching has been defined as an assessment of individual characteristics in relation to job requirements. The purpose of this study was to assess the inter-rater reliability of a job matching assessment instrument designed for young adults with disabilities. Participants included 21 young adults with intellectual disabilities. Post-high school teachers and job coaches served as raters. The young adult participants first identified 2–3 preferred jobs. Five pairs of raters completed the job matching assessment to identify the job that was best matched to the participant’s current characteristics. Ratings were based on 106 O*NET job dimensions related to knowledge, skill, and ability. Findings indicated relatively high inter-rater agreement. However, across job dimensions, data indicated high variability with disagreements apparent for particular ones (e.g., Knowledge of English Language, Safety Awareness Skill). Findings are discussed in relation to the importance of identifying preferred, well-matched jobs for individuals with intellectual disabilities.

Keywords: Job matching, job preference, web-based job assistance and accommodations, occupational information network (O*NET), inter-rater reliability, job dimensions, index score differences, extreme ratings

1. Introduction

For young adults with intellectual disabilities to be successful in supported employment, numerous factors must be considered. First, supports in the form of job coaching and co-worker assistance must be in place [5, 8]. Second, situational assessment and on-the-job training must be provided by personnel who, at least initially, have frequent contact with the employee [4]. Third, rehabilitation counselors, employers, family, and other stakeholders must play active roles in offering assistance, such as accommodations, transportation, and job modification [7]. Among other factors, two may need to be considered in the pre-placement stage: Identification of self-determined job preferences and degree of match to one’s preferred jobs. Job preference has been defined as responses by individuals indicating a choice of one stimulus (in this case, a representation of an occupation) over another [2]. Identification of job preferences is consistent with Title I, Part A, General Provisions of the Rehabilitation Act of 1998, as well as assessment involving interests and preferences required by special education legislation, i.e., Individuals with Disabilities Education Act (IDEA: 2004). Job matching has been defined as an assessment of individual characteristics in relation to job requirements [3, 11]. Swenson [10] referred to job matching as the extent to which an individual’s cognitive abilities, interests, and personality traits matched requirements in a particular job. Although web-based job matching services exist
Stakeholders need an instrument to assess the degree of match between preferred jobs, skills of job seekers, and job requirements. If followed by situational assessment to verify job match, such an instrument would increase the probability of successful job placement. One Internet-based site assesses both job preference and degree of match (www.yesjobsearch.com). The job seeker works at a computer making decisions about preferred work by watching videos of actual jobs. A narrator describes critical job tasks. After watching a video, the job seeker decides to keep or discard the job. The result is a list of 2–3 highly preferred jobs. Preference assessment takes the job seeker about 45–60 min to complete. A rehabilitation counselor remains available to answer questions. Thereafter, one or more stakeholders (such as a parent, transition teacher, or advocate) respond to the job matching assessment, i.e., a rating scale of 106 Occupational Information Network (O*NET: http://www.onetcenter.org/overview.html) dimensions describing knowledge, skills, and abilities (e.g., time-telling skills, knowledge of English language). Job matching assessment requires the stakeholder about 20 min to complete.

The O*NET data were generated by researchers [9] under contract with O*NET to measure the importance of various types of knowledge, skills, and abilities to specific occupations. Surveys of numerous occupations produced importance ratings for specific occupations. Workers and job analysts were presented with definitions of knowledge, skills, and abilities (i.e., dimensions) and asked to rate the importance of the dimension was required for a job. A mean of 18.5 workers and 6.0 job analysts rated each dimension. Inter-rater reliability among workers for skills ranged from 0.71 to 0.93, and for knowledge ranged from 0.37 to 0.95. Researchers [9] did not collect data on whether dimensions were usable for job matching for individuals with disabilities.

These data were placed in the O*NET database and were imported, given proper acknowledgement, to the www.yesjobsearch website. O*NET data were compiled on the relative importance of various areas of knowledge, skills, and abilities to each job. Next, the extent to which an individual job seeker with intellectual disabilities displays various areas of knowledge, skills, and abilities is determined by stakeholder ratings in the www.yesjobsearch assessment.

On the website, a computer program multiplies the O*NET importance weighting for a particular job by the rater’s rating of the job seeker’s skill levels (Good,
Fair, or Poor) for each of the 106 dimensions. The output is an index score on the extent to which the job seeker’s current skill level matches the job requirements for preferred jobs. The program also generates a list of strengths and weaknesses that may be used by stakeholders for training or accommodation.

Although this website provides information to stakeholders on the degree of match between preferred jobs, skills, and job requirements, the job matching instrument has not been subjected to psychometric assessment, such as inter-rater reliability. It is unknown how consistent ratings may be across raters. Consistent ratings may increase credibility regarding job prospects, whereas inconsistent ratings may result in misinformed decisions regarding training priorities, job accommodations, or job placement. The purpose of the current study is to assess the inter-rater reliability of the job matching assessment.

2. Method

2.1. Participants and settings

Participants included both young adults with intellectual disabilities and raters. Twenty-one young adults with intellectual disabilities participated. The mean age of the young adults was 19 years, 8 months. Overall, their mean IQ score was 59 (standard deviation = 4.8). Their job matching data were generated by five pairs of raters. Each pair consisted of a teacher and a job coach. Teachers were licensed special education professionals working in a post-high school program in a Western U.S. State. The mean age of teachers was 37 years, 4 months. Mean teaching experience was 8.4 years. Job coaches were paraprofessionals working under supervision of the teacher (the paired rater). While teachers occasionally observed young adults in community employment settings, job coaches supervised young adults on a daily basis. The mean age of job coaches was 30 years, 7 months. Mean job coaching experience was 3 years, 7 months.

Settings for the study were two post-high school transition programs. Post-high schools provide educational and employment training for young adults ages 18 through 21 when they remain eligible for special education services. One post-high program was sponsored by a large suburban school district and supplied 18 young adult participants and four pairs of teachers/job coaches. The second post-high program, sponsored by two school districts, was located on a university campus and supplied three young adult participants and one teacher/job coach pair. Both post-high school programs were committed to finding preferred, well-matched employment for young adults.

2.2. Job matching assessment and procedures

Teachers and job coaches independently completed the job matching assessment. Teachers and job coaches were explicitly told not to communicate with each other prior to or during the assessment regarding a young adult’s knowledge, skills, and abilities; ratings of particular items; or assessment results. The post-high school director supervised teachers and job coaches as they completed the assessment by either directly observing the rater completing the instrument or making arrangements for raters to work independently in vacant offices or work areas. Raters received the following printed instructions:

Step 1: Review the list of Job Dimensions below (e.g., Active Learning/Active Listening). See the Description for more information.

Step 2: Rate the participant’s current behavior or performance on each job dimension. Rate the participant by circling Good, Fair, or Poor. For example, if the participant has fair knowledge of the dimension named Mathematics Skills (see definition below), circle Fair. Rate each job dimension by circling Good, Fair, or Poor. Do not skip any. If you don’t know, circle DK.

Figure 1 shows the first seven of 106 dimensions. Figure 2 lists all 106 dimensions but, for parsimony, corresponding definitions are not provided. Definitions are available from the researcher. Dimensions and definitions were similar to knowledge, skill, and ability dimensions published by O*NET, Version 9 in 2007. The researcher modified some definitions to simplify language. Raters responded to each of 106 job dimensions, and were told by the researcher to base their ratings on the extent to which the young adult was prepared to independently perform tasks related to employment.

2.3. Measures

2.3.1. Familiarity of raters with young adult participants

Prior to rating skills, knowledge, and abilities of participants, raters were asked two questions to assess
their familiarity with the young adults participants: (a) Approximately how long ago (in days) did you last observe this participant’s community job skills? (b) Approximately how long (in total duration: Hours, days, months) have you supervised this participant’s community job skills? These questions were presented to obtain information on how often and how recent the rater had observed the participant’s skills, knowledge, and abilities.

2.3.2. Inter-rater agreement

The researcher created paper versions of the web-based job matching assessment to facilitate the two-rater rating process. Programming of the web site allowed for only one rater’s data to be stored; data from multiple raters could not be stored. Therefore, paper versions of the assessment were used to measure agreement between two independent raters. The paper versions were identical to the web-based assessment in terms of item sequence, rating procedures, and instructions. Raters received the assessment from the post-high school director, completed the ratings in their private offices or work areas, and returned the completed assessment to the director. The researcher collected completed assessments from the director and conducted scoring procedures. Only identical ratings across raters for a job dimension (e.g., Rater A = Good, Rater B = Good) were scored as agreements. Disagreements between raters could take two forms, e.g., Rater A = Good, Rater B = Fair or Rater A = Good, Rater B = Poor. The researcher recorded agreements and disagreements according to dimension number, extreme ratings (i.e., good/poor combinations), and dimension numbers with “Don’t Know” responses. If a rater responded “Don’t Know” on a dimension, it was discarded and not scored for inter-rater agreement. “Don’t Know” responses were tallied for pairs of raters. Agreements and disagreements were adjusted based on total “Don’t Know” responses. For example, if there were 70 agreements, 30 disagreements, and six “Don’t Know” responses, agreement and disagreement scores were adjusted by multiplying by the proportion of “Don’t Know” responses to the total items (e.g., 6 Don’t Know out of 106 = 0.0566, 70 × 1.0566 = 73.96 (rounded to 74 agreements), 30 × 1.0566 = 31.69 (rounded to 32 disagreements)).

Prior to administering the job matching assessment, the researcher worked individually with participants to administer the job preference assessment. The

<table>
<thead>
<tr>
<th>Rating</th>
<th>Job Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>1. Active Learning/Active Listening</td>
<td>Understanding the implications of new information for problem solving and decision making. This includes giving full attention to what other people are saying, taking time to understand, asking questions as appropriate, and not interrupting at inappropriate times.</td>
</tr>
<tr>
<td>Good</td>
<td>2. Learning Strategies</td>
<td>Selecting and using training methods and procedures for the situations when learning or teaching new things.</td>
</tr>
<tr>
<td>Good</td>
<td>3. Mathematic Skills</td>
<td>Using mathematic skills to solve problems.</td>
</tr>
<tr>
<td>Good</td>
<td>4. Money Skills</td>
<td>Ability to accurately add, subtract, multiply, and divide amounts of money. Also, ability to accurately compare sums of money, use basic banking facilities, and independently make informed, appropriate decisions regarding money.</td>
</tr>
<tr>
<td>Good</td>
<td>5. Time-Telling Skills</td>
<td>Ability to independently report the correct time, add time (example: “in two hours from now when it is 10:30”), report when events will take place, and estimate the time necessary to complete tasks.</td>
</tr>
<tr>
<td>Good</td>
<td>6. Reading Comprehension</td>
<td>Understanding written sentences and paragraphs in work-related documents.</td>
</tr>
<tr>
<td>Good</td>
<td>7. Written Comprehension and Expression</td>
<td>The ability to read and understand information and ideas presented in writing. The ability to communicate information in writing, keyboard, or other format so others will understand.</td>
</tr>
</tbody>
</table>
1. Active Learning/Active Listening  
2. Learning Strategies  
3. Mathematic Skills  
4. Money Skills  
5. Time-Telling Skills  
6. Reading Comprehension  
7. Written Comprehension and Expression  
8. Job-Related Reading Skills  
9. Speaking  
10. Speech Clarity  
11. Speech Recognition  
12. Oral Comprehension and Oral Expression  
13. Writing  
14. Time Management  
15. Time Sharing  
16. Social Coordination  
17. Personnel and Human Resources  
18. Custodianship and Personal Service  
19. Instinct  
20. Negotiation  
21. Persuasion  
22. Service Orientation  
23. Monitoring  
24. Judgment and Decision Making  
25. Deductive Reasoning  
26. Inductive Reasoning  
27. Critical Thinking  
28. Flexibility and Speed of Closure  
29. Fluency of Ideas  
30. Equipment Maintenance/Equipment Selection  
31. Installation  
32. Controlling, Monitoring, and Analyzing Operations  
33. Programming  
34. Quality Control Analysis  
35. Repairing  
36. Technology Design  
37. Troubleshooting  
38. Memorization  
39. Originality  
40. Visualization  
41. Problem Sensitivity  
42. Selective Attention  
43. Perceptual Speed  
44. Spatial Orientation  
45. Dynamic Flexibility  
46. Dynamic Strength  
47. Explosive Strength  
48. Gross Body Coordination  
49. Gross Body Equilibrium  
50. Stamina  
51. Static Strength and Trunk Strength  
52. Arm-Hand steadiness  
53. Control Precision  
54. Finger Dexterity and Manual Dexterity  
55. Multi-limb Coordination  
56. Rate Control  
57. Reaction Time  
58. Response Orientation  
59. Speed of Limb Movement  
60. Wrist-Finger Speed  
61. Auditory Attention  
62. Hearing Sensitivity  
63. Sound Localization  
64. Depth Perception  
65. Far Vision  
66. Near Vision  
67. Night Vision  
68. Peripheral Vision  
69. Glare Sensitivity  
70. Visual Color Discrimination  
71. Administration and Management  
72. Economics and Accounting

Fig. 2. List of 106 job matching dimensions.
researcher communicated to raters the three jobs identified as highly preferred. Therefore, raters were aware of the participant’s three highest preference jobs as they conducted the job matching assessment.

2.3.3. Index score differences

As mentioned previously, the output of the job matching assessment was an index score on the extent to which the job seeker’s current skill level matched the job requirements for preferred jobs. Index scores ranged from 0.00 to 1.00. A high score of 0.75 or above indicated the job was within the participant’s current knowledge, skills, and abilities to perform, although training may still be necessary to meet standards and formal requirements of the job. A low score of 0.50 or below indicated the job may require considerable training or accommodation for the participant. These cut-off scores were established based on previous research [6]. The researcher entered rating data from each rater into the web-based job matching assessment to obtain the index score. Index score differences provided a descriptive measure of the variability between raters.

2.3.4. Extreme ratings (i.e., good/poor combinations)

The researcher tallied “Good/Poor” ratings on specific dimensions to detect extreme differences. For example, if one rater rated Reading Comprehension as “Poor” and the other rater rated the same item as “Good”, the combination was recorded. Total extreme ratings were recorded for rater pairs and for specific dimensions within the assessment.

2.3.5. Data analysis

Data on inter-rater agreement were expressed by computing a mean agreement score out of 106 items across rater pairs. Familiarity data were expressed as months elapsed since last observation and total duration of supervision of a participant’s community job skills. Index difference scores were expressed as number of points difference across raters (e.g., 5 points: 0.41 for Teacher and 0.36 for Job Coach) for each participant. Extreme ratings were tallied and the mean number for all rater pairs was calculated. Additionally, extreme ratings were tallied according to job dimension.

<table>
<thead>
<tr>
<th>73. Biology and Chemistry</th>
<th>74. Physics</th>
<th>75. Medicine and Dentistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>76. Science</td>
<td>77. Building and Construction</td>
<td>78. Engineering and Technology/Mechanical</td>
</tr>
<tr>
<td>79. Design</td>
<td>80. Clerical</td>
<td>81. Communications and Media</td>
</tr>
<tr>
<td>82. Computers and Electronics</td>
<td>83. Telecommunications</td>
<td>84. English Language</td>
</tr>
<tr>
<td>85. Foreign Language</td>
<td>86. Geography</td>
<td>87. Law and Government</td>
</tr>
<tr>
<td>88. Mathematics Knowledge</td>
<td>89. Education and Training</td>
<td>90. Psychology, Sociology, and Service</td>
</tr>
<tr>
<td>91. Fine Arts</td>
<td>92. Public Safety and Security</td>
<td>93. Sales and Marketing</td>
</tr>
<tr>
<td>94. Transportation</td>
<td>95. Food Production</td>
<td>96. Production and Processing</td>
</tr>
<tr>
<td>97. Appearance and Hygiene</td>
<td>98. Appropriate Communications to Co-Workers and Employer</td>
<td>99. Appropriate Communications to Public</td>
</tr>
<tr>
<td>100. Following Directions</td>
<td>101. Responding Appropriately to Correction</td>
<td>102. Recognizing and Reporting Potential Hazards/Emergencies</td>
</tr>
<tr>
<td>103. Responding to Hazards/Emergencies</td>
<td>104. Safety Awareness</td>
<td>105. Punctuality</td>
</tr>
<tr>
<td>106. Initiative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. (Continued)
3. Results

3.1. Inter-rater agreement

Figure 3 shows the distribution of inter-rater agreement data out of 106 job dimensions. All data are adjusted to account for “Don’t Know” responses. Mean agreement across raters was 76.05 agreements out of 106 dimensions (71.75%). Therefore, mean disagreements were 29.95 out of 106 dimensions (28.25%). Agreements ranged widely from 64 to 89 out of 106 (standard deviation = 6.26). Inter-rater data for Teacher-Job coach pairs is shown in Table 1. A chi-square test for differences in a proportion [1] was computed to determine whether mean agreement was greater than what might be expected by chance. Because disagreements could take two forms (e.g., Rater A = Good, Rater B = Fair; Rater A = Good, Rater B = Poor) and agreements could take only one form (e.g., Rater A = Good, Rater B = Good), expected agreement was 0.33. The mean observed agreement (0.71) was compared to expected agreement (0.33). The difference was statistically significant (chi square = 3.78, \( p < 0.005 \), df = 20).

“Don’t Know” responses ranged from zero to 20 out of 106 (mean = 3.81). Items with most common agreement were those related to academic knowledge (e.g., Biology, Chemistry, Physics). These dimensions were commonly rated “Poor”. Mean agreement data were examined to detect differences across rater pairs, who rated between two to seven participants. Mean agreement across pairs of raters ranged from 72.40 to 79.71 (standard deviation = 3.52). There were no statistically significant differences in rating scores between raters or across pairs of raters. Across 21 young adults with disabilities, best-matched and highest preference jobs were Dining Room Attendant (n = 5), Theatre Worker (n = 4), and Greenhouse Worker (n = 3).

3.2. Familiarity

Familiarity questions were presented to obtain information on the recency of supervision and familiarity of the rater with the participant’s skills, knowledge, and abilities. When asked “how long ago did you last observe this participant’s community job skills”, teachers’ mean response was about 24 days and job coaches’ mean response was about 6 days. These data indicate that, based on self-report, both teachers and job coaches were familiar with participant skills based on direct observation, although job coaches had observed participant performance more recently. When asked “how long (in hours, days, or months) have you supervised this student’s community job skills”, teachers’ mean response was 7.95 months and job coaches’ mean response was 10.76 months. These data indicate that job coaches had supervised participants slightly longer than teachers, but each for sufficient duration to rate knowledge, skills, and abilities.

3.3. Index difference scores

The mean difference in index score across raters was 0.0929 points (range = 0.000 to 0.190 points, standard deviation = 0.0426). For one participant, a pair of raters had an identical index score. For six participants, pairs of raters had index difference scores within 0.05 points. On the other extreme, for eight participants, pairs had index difference scores of 0.130 points or higher. Index scores ranged from 0.87 to 0.28 (mean = 0.54).

3.4. Extreme ratings

The mean number of extreme ratings for all rater pairs was 1.76 (range = 0 to 5) out of 106 dimension ratings. Job dimensions with most extreme ratings included Knowledge of English Language (n = 3 out of 21 participants) and Safety Awareness Skill (n = 3 out of 21 participants).
The assessment received extreme ratings from at least one pair of raters.

4. Discussion

Results of this study provided data on the inter-rater reliability of a job matching assessment for 21 young adults with intellectual disabilities. Data indicate relatively high inter-rater agreement and low index score differences. Mean agreement between raters was 71.4% when the probability of agreement was 33.33%; therefore, agreement was significantly higher than that expected by chance. However, results should be interpreted cautiously with particular attention to two dimensions. First, results show high variability between raters for specific individuals. For three participants, pairs of raters agreed on less than 66.67% (70 or less) of 106 ratings. In many cases where raters disagreed, comparison of ratings revealed patterns. For example, Rater A would rate “Good” while Rater B would rate “Fair”, then when Rater A shifted to “Fair” on a subsequent item, Rater B shifted to “Poor”. Despite patterns between raters, this tendency speaks to the need for more specific definitions of dimensions and clarification of expectations of raters. Second, raters rated a mean of between one and two dimensions with extreme scores out of 106 dimensions, and interestingly, the most common dimensions with high extreme scores are relatively important to many jobs (Knowledge of English Language, Safety Awareness Skills). Extreme scores decreased overall index score agreement for individual participants. More important, even occasional extreme scores are difficult to explain and limit the interpretation of data produced by the assessment. At minimum, they echo the need for specific definitions of dimensions.

Relatively high variability across pairs of raters and occasional occurrence of extreme ratings on specific job dimensions signal the need to examine the instrument itself for potential revision. Two questions arise. First, can better clarity be achieved in the instructions? Raters were told to base their ratings on “the extent to which the young adult was prepared to independently perform tasks related to employment at the present time”. Such a statement may have been vague and subjective because “employment at the present time” did not specify a particular context (e.g., types of skills or employment). Second, were the names and definitions of certain dimensions unclear, thus resulting in different ratings across raters? In regards to knowledge of English language, safety awareness skills, and perhaps other dimensions, raters may have interpreted definitions differently or their bearing in criteria for employment. A follow-up debriefing with participants may be useful in revealing definitions requiring greater clarification.

In addition to issues mentioned above, three weaknesses are noteworthy and limit the generalizability of study results. First, a relatively low number of participants were involved, particularly in regards to pairs of raters (n = 5). Additional research is needed with larger numbers of raters as well as an increased sample of job seekers with disabilities. Second, the present study used paper versions of the assessment as opposed to the existing web-based version for programming reasons described previously. It is unknown whether responses to the web-based assessment would have produced identical ratings by a particular rater. Although programming of the web site prevented multiple raters, it might be worth revisiting this issue and determining the feasibility of multiple raters. Although high levels of inter-rater reliability are desirable, job-matching assessments should accommodate multiple raters to increase credibility or to explore differences across raters. Third, although recency data provide index of how long it had been since teachers and job coaches observed performance, they provided no information about what job was observed. Observations may/may not have been related to a high preference job, which may have affected performance and the observer’s evaluation. In turn, a recent observation of poor performance on a low preference job may have affected the rater’s judgment on particular dimensions (e.g., judgment and decision making, selective attention, stamina, etc.).

Despite the limitations of this study, these findings appear to be the first to examine job matching along with job preference assessment for individuals with intellectual disabilities. The concept of assessment to determine well-matched, high preference jobs seems conceptually reasonable as a precursor to job placement. While job-matching services exist for job seekers in the corporate world, such an assessment takes on heightened priority for individuals with intellectual disabilities for three reasons. First, their skills vary widely so degree of match across a variety of jobs is likely to vary. Second, successful job matching for individuals with disabilities takes on paramount importance. Unsuccessful matches may tarnish an individual’s motivation to participate in the workforce. Similarly,
unsuccessful matches may diminish an employer’s motivation to repeat the process of hiring job seekers with intellectual disabilities. Third, the absence of job matching assessment leaves the judgment of “goodness of fit” to stakeholders who have only their subjective point of view to lend to the process. Even when stakeholders discuss the degree of match among different jobs, they may not produce a job placement decision with a high level of reliability.

To the rehabilitation counselor, job preference and matching data on an individual job seeker with intellectual disabilities may be cost efficient and useful. Cost efficiency is evident because the rehabilitation counselor’s time is not consumed in assessment. The job seeker spends 45–60 min identifying preferences and a stakeholder spends 20 min completing the job matching assessment. Although the rehabilitation counselor must remain available to answer questions, she/he may carry out other tasks. The rehabilitation counselor reviews printouts from both assessments and can immediately undertake the job development process. Usefulness is evident because an individual job seeker’s preferences are identified from the outset and initiate the job placement process. Additionally, the job matching index score conveys to the rehabilitation counselor the extent to which current skills match requirements in the preferred jobs.

Future research should examine the relationship between degree of match predicts successful placement (e.g., length of employment, employee and employer job satisfaction). It may be possible to predict the probability of high productivity, satisfaction, retention, and other variables given job match index scores and placement in high preference jobs.

References
