Master Tutorial

TITLE

Web Scraping with R

SHORTENED TITLE

Web Scraping with R

ABSTRACT

A lot of data exists on the web! Accessing that data requires understanding HTTP requests, security tokens, data transfer file formats, and data cleaning. This tutorial session will walk you through how to access and process web-based data using the powerful statistical language R. Bring your laptop (optional) for this interactive session (download session materials here: ).

PRESS PARAGRAPH

Gathering data no longer requires handing out questionnaires or setting up experiments. Many interesting questions can be answered by accessing data on the web. The statistical computer language R allows scientists and practitioners to scrape the web and obtain a world of data without leaving one’s seat, including the distribution of the population, the sentiment of consumers and employees, and the change in stock prices. This session will give a hands-on and step-by-step tutorial on how to access and process web-based data for the purposes of analyses.

WORD COUNT: 1840

**Web Scraping with R**

R (R Core Team, 2015) is an open-source programming language that is designed for statistical computing (Hornik, 2013). R can perform anything from standard data analysis (e.g. Multiple Regression, Hierarchical Linear Modeling, or Structural Equation Modeling) to highly specialized computations that may be unique to a scientific field. R is a programming *language* and not just a statistical analysis package. Many data scientists and practitioners can contribute to R by writing new and unique software, called “packages” in R. At the time of this writing, the Comprehensive R Archive Network (CRAN) contains 12,959 available packages, including packages to read data in varying formats (e.g., readr, open.xlsx, haven, rjson, officer), access databases (e.g., DBI, odbc, RSQLite), clean data (e.g., dplyr, tidyr, stringr, reshape2), perform data analyses and machine learning (e.g., caret, xgboost, randomForest, caret, survival), visualize results (e.g., shiny, ggplot2), and interface with other programming languages (e.g., Rcpp, reticulate, RJava). These packages, just like R itself, are free of charge.

This Master Tutorial will teach attendees how to leverage R and several packages to access data from various websites and put that data into a form useful for data analysis. We will teach users how to request information from servers, access web APIs (including those that require authentication keys), extract useful information from web requests, and process online data for the purposes of further analyses. Attendees should be familiar with R and have both R and RStudio installed prior to the workshop. We will walk through and explain each line of code in detail, but we will have little time to review the basics of R itself. # TO HERE #

*Motivation for Session*

The proposed tutorial is a continuation of the authors’ ongoing attempt to make R more accessible to the I/O community and lower the entry barrier for scientists and practitioners. R’s user base is quickly increasing. R usage is on par with or has surpassed the most popular packages like SPSS and SAS in the academic realm (Muenchen, 2015). R’s increase in popularity may be best described by the ecosystem of support pages, books, blogs, tutorials, and even R specific conferences and journals that have developed around this programming language. R has recently been ranked as fifth among the top ten programming languages (Smith, 2016), and recent polls (Piatetsky, 2015) indicated that R is the most popular analysis software among data scientists. In the I/O community R is on the rise as well. Topics such as Big Data and Machine Learning are gaining attention in conference submissions and journals, and R is frequently used for these purposes (Jones, Goebl, & Semmel, 2016). Building on two preceding SIOP conference contributions (Schwall, Lustenberger, Beatty, & Jones, 2014; Schwall, Beatty, & Jones, 2015) this session will aim at teaching how to put powerful tools into the hands of more I/O psychologists. Below is a summary of the two packages that will be subject of this tutorial:

Package **ggplot2** (Wickham, 2009) is a powerful data visualization tool that moves far beyond the base R plotting capabilities. It both simplifies the creation of basic, but professional, graphs, and allows for extreme customization and layering for more advanced and specialized purposes. Package ggplot2 allows users to create a vast variety of graphs ranging from simple histograms, scatter plots, bar plots, to dendrograms, heat maps, three dimensional scatter plots, lattice plots, vector fields, mosaic plots, and geographic maps. Given its versatility, ggplot2 is a critical tool for any type of data visualization task.

Package **shiny** (Chang, Cheng, Allaire, Xie, & McPherson, 2015) allows the display of interactive graphs and plots (e.g. plots in ggplot2) on websites. These websites are highly customizable. shiny’s purpose is to allow scientists and practitioners to easily communicate their findings by publishing them on the internet without having to use HTML, JavaScript, and server-side languages (C#, Python, Ruby, etc.). The user creates an R script which in turn generates the required HTML and JavaScript. Excellent interactive examples are housed here: <http://shiny.rstudio.com/gallery/> and here: <https://www.rstudio.com/products/shiny/shiny-user-showcase/>

*Presenters*

Each of the presenters of this tutorial has trained other R users in the past in the use of base R and the use of the packages in question. In addition, they all have been in the situation of being a brand new R user of these packages and they do not have a programming or computer science background. They are keenly aware of the obstacles that R poses specifically to I/O psychologists without background (or interest) in programming and they have found methods of teaching R both internally in their organizations and as part of the SIOP community. Adam Beatty has used R to conduct analyses for a number of published articles and is viewed as an organizational resource on R at HumRRO. Jeff Jones has published several R routines in methodological and theoretical journals such as *Psychometrika* and *Psychological Methods*, and trains I/O psychologists in R at Korn Ferry. Alexander Schwall is leading the organization-wide implementation of R as a central analysis instrument at DDI and is training DDI associates in R. All three are recognized R resources and R power users in their organizations. More details on the three presenters will be provided in the Appendix of this document.

*Tutorial Plan*

This group of presenters led a panel discussion at the 2014 SIOP conference and a Master Tutorial at the 2015 SIOP conference. Both sessions were very well attended. With this tutorial we want to offer an additional opportunity to get started with R. This session will provide simple and easy to follow instructions that the audience can either replicate on their own computer during the session or simply follow along. The proposed master-tutorial will provide interactive instruction to address the following three learning objectives:

1. Opening R and preparing the data
2. Exploring and Visualizing data with ggplot2
3. Creating web-based (browser-based) graphs and figures with shiny.

These three learning objectives will give the audience members first basic steps to understand the functionality of these packages, create basic example applications, and to use these example applications as templates to create their own working R scripts in the future.

Note that the tutorial will be interactive. The audience is encouraged to follow the programming steps on their own laptops, however, this will *not be required* to effectively learn in this tutorial. Interested attendees will be encouraged in advance of the conference to download the R software and packages. Materials for the tutorial will be provided using either SIOP’s mySIOP.org file repository or a link to a file repository from which the users can download the required files <http://bit.ly/1UyrpYw>.[[1]](#footnote-1) Moreover, the shiny dashboard that will be created during the session will be hosted ahead of time on the site <https://siop2017.shinyapps.io/Tutorial/> for interested parties to explore.

The tutorial will follow the following general process: The presenter will explain and demonstrate each step in R, and then give the audience the opportunity to perform this step by themselves on their own computer. Through this process the audience members will assemble their own example R script that they will be able to use in the future as a starting point for other projects. While one presenter will lead a segment, the other two will be addressing audience help requests. Audience members can also choose to open a completed syntax script that they can run step by step. This script will contain the materials discussed in the tutorial but also additional elements that the audience members can work through in self-study after the tutorial. We want to ensure that motivated audience members could use the script as a template to start their own work and to gain further proficiency.

Time requested for this tutorial is 80 minutes. The approximate time for each learning objective is noted below, along with further descriptions. Note that none of the authors are affiliated with the producers of the packages and that there are no material gains (financial or otherwise) for them. All software packages presented are full versions (as opposed to trial versions) and free of cost. Our main goal is to *raise awareness* and provide basic skills for these packages so that the I/O community does not miss out on the capabilities of these powerful tools.

**Learning Objective #1: Opening R and preparing the data (20 minutes)**

As a preliminary step, the authors will show how to start R, familiarize the audience with the software, and show how to read data into R. For this tutorial we will use RStudio which makes many frequently performed tasks quite easy and user friendly. Given that one’s data is often not in readily usable format we will then spend also spend a few minutes showing how to shape and wrangle datainto the desired format. The goal is that even completely inexperienced users could load their data and prepare it for further processing.

**Learning Objective #2: Exploring and visualizing data with ggplot2 (25 minutes)**

We will start by building a simple first graph, explaining the elements that form the basic structure for this and any graph in ggplot2. Then, step by step, we will customize this graph by changing elements of the script and showing the results of these changes. All of these changes can be executed by the audience in a script that is made available in the download folder (see above). From there, we will branch out into showing how to convert the current type of graph into other types (e.g. we will convert a scatter plot to a line plot). ggplot2 is known to be a daunting for first time users as many new commands need to be learned. We will overcome this hurdle by using a gentle step-by-step approach in which we introduce commands one at a time. Finally, we will demonstrate how to change the visual appearance of the graph (e.g. background, colors, labels, tick marks) through the themes function.

**Learning Objective #3: Creating web-based (browser-based) graphs and figures (25 minutes)**

In this section, we will demonstrate how to setup an R shiny application. Specifically, we will build upon the previous sections and integrate a ggplot2 plot example into a working, interactive, web-deployable dashboard. Moreover, we also show how to take the newly created interactive dashboard and host it on the web via a free account on <http://www.shinyapps.io/>. This step-by-step approach will ensure that the audience will not get overwhelmed by the numerous commands that need to be written. As a result, this section will provide a template that will allow users to present graphs, figures, and text on the web. In essence, we will share the most common denominator for any shiny page that the audience members then can use for their own work. This session will get users “over the hump” by providing a template that can be filled in by the user.

**Conclusion and Summary (10 minutes)**

The presenter will try to leave a 10 minute buffer at the end of the session. This 10 minute buffer will be used to answer audience questions and help with technical problems. The presenters will point participants to sections in the syntax script that could serve as good starting points for further self-study and share print-outs with helpful resources for R beginners.

**Presenter Information**

Adam Beatty

Research Scientist

Human Resources Research Organization (HumRRO)

100 Washington Avenue, Suite 1660

Minneapolis, MN 55401

Phone: 612-486-1026

Email: [abeatty@humrro.org](mailto:abeatty@humrro.org)

Membership Status: Member

Jeff Jones

Sr. Manager, Analytics

Korn Ferry

33 South Sixth Street

Suite 4900

Minneapolis, MN 55402

Phone: 612-373-3418

Email: [jeff.jones@kornferry.com](mailto:jeff.jones@kornferry.com)

Membership Status: Member

Alexander Schwall

Sr. Consultant

Development Dimensions International, Inc.

1225 Washington Pike

Bridgeville, PA 15017

Phone: 412-220-7937

Fax: 412-220-3599

Email: [Alexander.Schwall@ddiworld.com](mailto:Alexander.Schwall@ddiworld.com)

Membership Status: Member

**References**

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2015). shiny: Web Application Framework for R. R package version 0.12.2. http://CRAN.R-project.org/package=shiny

Hornik, K. (2013). The R FAQ. Retrieved from [http://CRAN.R-project.org/doc/FAQ/R-FAQ.html](http://cran.r-project.org/doc/FAQ/R-FAQ.html)

Jones, J. A., Goebl, A. P., & Semmel, S. G. (2016, April). *Handling big(gish) data in R:*

*An introductory and interactive tutorial.* Master Tutorial at the annual meeting of the Society of Industrial and Organizational Psychology, Anaheim, CA.

Muenchen, R. (2015). The Popularity of Data Analysis Software | r4stats.com. Retrieved September 1, 2015, from http://r4stats.com/articles/popularity/

Piatetsky, G. (2015). R leads RapidMiner, Python catches up, Big Data tools grow, Spark ignites. Retrieved August 29, 2015, from http://www.kdnuggets.com/2015/05/poll-r-rapidminer-python-big-data-spark.html#prog.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

R Markdown — Dynamic documents for R. (n. d.). Retrieved from rmarkdown.rstudio.com

RStudio - RStudio IDE. (n.d.). Retrieved September 2, 2015, from http://www.rstudio.com/ide/

Schwall, A., Lustenberger, D., Beatty, A., & Jones, J. A. (2014, May). *Getting started*

*with R: Examples and Lessons Learned.* Panel discussion at the annual meeting

of the Society of Industrial and Organizational Psychology, Honolulu, HI.

Schwall, A., Beatty, A., & Jones, J. A. (2015, May). *Getting started with R: An*

*Interactive and Hands on Tutorial.* Master Tutorial at the annual meeting of the

Society of Industrial and Organizational Psychology, Philadelphia, PA.

Smith, D. (2016). R moves up to 5th place in IEEE language rankings. Retrieved August 10, 2016, from<https://www.r-bloggers.com/r-moves-up-to-5th-place-in-ieee-language-rankings/>.

The R Journal. (n.d.). Retrieved September 5, 2013, from http://journal.r-project.org/index.html

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.

Xie, Y. (2015). knitr: A General-purpose package for dynamic report generation in R. R package version 1.11.

**Appendix**

CV Adam Beatty

CV Jeff Jones

CV Alexander Schwall

**ADAM S. BEATTY**

100 Washington Avenue S, Suite 1660

Minneapolis, MN 55401

Phone: (612) 486-1026 (work)

Email: abeatty@humrro.org

***EDUCATION***

|  |
| --- |
|  |

**December 2013** *Ph.D., Industrial/Organizational Psychology*

University of Minnesota - Twin Cities

**May 2005** *Bachelor of Arts (Psychology),* *Magna Cum Laude*

St. John's University, Collegeville, MN

***WORK EXPERIENCE***

|  |
| --- |
|  |

**2011 – Present** *Research Scientist, Human Resources Research Organization*

**2007-2013** *Research Assistant, University of Minnesota*

**2006-2012** *Psychometrist, MDA Leadership Consulting*

***AWARDS AND HONORS***

|  |
| --- |
|  |

**2014; 2015 Lola Zook Award for Excellence in Technical Communications**

**2011-2012 HumRRO Meredith P. Crawford Fellowship**

A one-time competitive award provided to a dissertation-stage I/O

psychology graduate student demonstrating exceptional research promise.

**2009; 2010; 2013** **Society for Industrial and Organizational Psychology (SIOP) Featured Top Poster**

**2010; 2012** **Graduate Summer Research (GSR) Grant**

*University of Minnesota, Twin Cities*

**2009-2012** **CLA student Travel and Research Award (won each year)**

*University of Minnesota, Twin Cities*

**2010**  **Robert W. Eichinger Graduate Fellowship**

*University of Minnesota, Twin Cities*

**2009**  **Graduate Research Partnership Program (GRPP) Summer Grant**

*University of Minnesota, Twin Cities*

**2007-2008 Graduate School Fellowship**

*University of Minnesota, Twin Cities*

***PROFESSIONAL AFFILIATIONS AND SERVICE***

|  |
| --- |
|  |

Member - Society for Industrial and Organizational Psychology (SIOP)

-Ad-hoc reviewer for annual SIOP conference (2011 – Present)

Minnesota Professional for Psychology Applied to Work (MPPAW)

*Journal of Occupational and Organizational Psychology* – Ad-hoc reviewer

*Educational Measurement: Issues and Practice* – Ad-hoc reviewer

*Personnel Psychology* – Ad-hoc reviewer

*Journal of Personnel Psychology* – Ad-hoc reviewer

***RESEARCH HISTORY***

|  |
| --- |
|  |

***Book and Journal Articles***

Beatty, A. S., & Walmsley, P. T. (2016). We found them! The practitioner role in expanding the generalizability of findings in I-O Psychology. *Industrial and Organizational Psychology, 9*, 181-187.

Sackett, P. R., Walmsley, P. T., Koch, A. J., Beatty, A. S., & Kuncel, N. R. (2016). Predictor content matters for knowledge testing: Evidence supporting content validation. *Human Performance, 29*, 54-71.

Higdem, J. L., Kostal, J. W., Kuncel, N. R., Sackett, P. R., Shen, W., Beatty, A. S., & Kiger, T. B. (2016). The role of socioeconomic status in SAT-freshman grade relationships across gender and racial subgroups. *Educational Measurement: Issues and Practice, 35*, 21-28*.*

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice, 34*, 31-40.

Huber, C. R., Kuncel, N. R., Sackett, P. R., & Beatty, A. S. (2015). Validity stability across entering college cohorts: Exploring the temporal generalizability of local validity estimates. *International Journal of Selection and Assessment, 23*, 237-246.

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How are we doing after 30 years? A meta-analytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management, 41,* 318-348.

Beatty, A. S., Barratt, C. L., Berry, C. M., & Sackett, P. R. (2014). Testing the generalizability of indirect range restriction corrections. *Journal of Applied Psychology, 99*, 587-598.

Beck, J. W., Beatty, A. S., & Sackett, P. R. (2014) On the Distribution of Job Performance: The Role of Measurement Characteristics in Observed Departures from Normality. *Personnel Psychology, 67*, 531-566.

Trippe, D. M., Moriarty, K. O., Russell, T. L., Carretta, T. R., & Beatty, A. S. (2014). Development of a cyber/information technology knowledge test for military enlisted technical training qualification. *Military Psychology, 26*, 182-198.

Kuncel, N. R., & Beatty, A. S. (2013). Measurement of intelligence: Consistent findings and new directions. In K. F. Geisinger (Ed.), *APA* *handbook of testing and assessment in psychology*. APA Handbooks in Psychology Series. Washington, DC: APA.

Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., Shen, W., & Kiger, T. B. (2012). The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychological Science, 23,* 1000-1007*.*

Shen, W., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., & Kiger, T. B. (2012). All validities are not created equal: Determinants of variation in SAT validity across schools. *Applied Measurement in* *Education, 25,* 197-219.

Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the "more is better" vs. "good enough" hypotheses. *Psychological* *Science, 22,* 1336-1342.

Beatty, A. S., Syzdek, M. R., & Bakkum, A. (2006). The Saint John's experience project: Challenging men's perceptions of normative gender role conflict. *The Journal of Men's Studies, 14*, 322-336.

***Technical Reports***

Putka, D. J., Paullin, C., Medsker, G. J., Beatty, A. S., Caramagno, J. P., & Ginter, R. (2014). *An evaluation of the value of HR certification for individuals and organizations* (2014 No. 078). Alexandria, VA: Human Resources Research Organization.

Trippe, D. M., Moriarty, K. O., Beatty, A. S., & Diaz, T. E. (2014). Cyber Test Form Development and Follow-on Cyber Applications (2014 No. 041). Alexandria, VA: Human Resources Research Organization.

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2013). *The reliability of college grades.* New York, NY: Research Report submitted to the College Board.

Koch, A. J., Kuncel, N. R., Sackett, P. R., Beatty, A. S., & Walmsley, P. T. (2013). *Exploring omitted variables: Towards an understanding of why cognitive tests overpredict minority performance.* New York, NY: Research Report submitted to the College Board.

Koch, A. J., Kuncel, N. R., Sackett, P. R., Beatty, A. S., & Walmsley, P. T. (2013). *Predicting undergraduates’ persistence in science, technology, engineering, and math majors.* New York, NY: Research Report submitted to the College Board.

O’Shea, P. G., Geimer, J. L., Paullin, C., Tsacoumis, S., Waters, S. D., Whetzel, D. L., & Beatty, A. S. (2013). *Support for the development of leadership assessments: high-potentials and current supervisor assessment: Final report* (2013 No. 080). Alexandria, VA: Human Resources Research Organization.

Putka, D. J., Fleisher, M. S., & Beatty, A. S. (2013). *Enhancing the predictive potential of personality: Isolating multiple components of trait expression via a single administration design.* (2013 No. 088). Alexandria, VA: Human Resources Research Organization.

Trippe, D. M., Beatty, A. S., & Swain, M. (2013). *2013 CRT scaling & equating.* (2013 No. 090). Alexandria, VA: Human Resources Research Organization.

Barnett, R. C., Bedford, C. L., & Beatty, A. S. (2012). *The validity of MDA’s comprehensive psychological assessment services.* Minneapolis, MN: MDA Leadership Consulting.

Beatty, A. S.,Sackett, P. R., Kuncel, N. R., Kiger, T. B., Rigdon, J. L., & Shen, W. (2012). *A comparison of alternate approaches to creating indices of academic rigor* (Research Report 2012-11). New York, NY: College Board.

O'Shea, G., & Beatty, A. S. (2012). Task and KSAO linkage study. In G. O’Shea, S. D. Waters, & C. Paullin (Eds.), *Analysis of system operator/power dispatching and natural gas control work* (FR-12-10; pp. 9–19). Alexandria, VA: Human Resources Research Organization.

Beatty, A. S.,Sackett, P. R., Kuncel, N. R., Kiger, T. B., Rigdon, J. L., & Shen, W. (2011). *Comparing the validity of the Academic Rigor Index to a set of empirical-weighting procedures*. Research Report submitted to theCollege Board, New York, NY.

Bynum, B.H., Koger, L. E., Beatty, A. S., Taylor, L. R., Deatz, R. C., Thacker, A. A. (2011). *Evaluation of the College Board Florida Partnership for Minority and Underrepresented Students: Year 2 report* (FR-11-48). Alexandria, VA: Human Resources Research Organization.

Rigdon, J. L., Shen, W., Kuncel, N. R., Sackett, P. R., Beatty, A. S., & Kiger, T. B. (2011). *The role of socio-economic status in SAT-freshman grade relationships across gender and racial subgroups.* Research report submitted to the College Board, New York, NY.

Shen, W., Kiger, T. B., Sackett, P. R., Kuncel, N. R., Walmsley, P., Beatty, A. S., & Rigdon, J. L. (2011). *Differential prediction in the use of the SAT and high school grades in predicting college performance: Joint effects of race and language*. Research report submitted to the College Board, New York, NY.

Trippe, D. M. & Beatty, A. S. (2011). CTN "A" School validation. In D. M. Trippe & T. R. Russell (Eds.), *Information and communication technology literacy test norming study: Phase III final report* (FR 11-47). Alexandria, VA: Human Resources Research Organization.

Trippe, D. M., Beatty, A. S., & Wise, L. L. (2011). Analysis of norming study data. In D. M. Trippe & T. R. Russell (Eds.), *Information and communication technology literacy test norming study: Phase III final report* (FR 11-47). Alexandria, VA: Human Resources Research Organization.

Trippe, D. M., Dickinson, E. R., Ford, L. & Beatty, A. S. (2011). 2011 *CRT scaling and equating* (FR-11-50). Alexandria, VA: Human Resources Research Organization.

Beatty, A. S., Sackett, P. R., Kuncel, N. R., Rigdon, J. L., Shen, W., & Kiger, T. B. (2010). *Testing the limits of empirically-based prediction of college freshman grade-point average and retention using information from the student descriptive questionnaire*. Research Report submitted to the College Board, New York,NY.

Shen, W., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., & Kiger, T. B. (2010). *All validities are not created equal: Determinants of variation in SAT validity across schools*. Research Report submitted to the College Board, New York, NY.

Sackett, P. R., Kuncel, N. R., Rigdon, J. L., Beatty, A. S., Shen, W., & Kiger, T. B. (2009). *Socio-economic status and the relationship between the SAT, high school grades, and freshmen GPA:* *An analysis of data from 110 colleges and universities using the 2005* *revision of the SAT*. Research report submitted to theCollege Board, New York, NY.

Shen, W., Beatty, A. S., & Sackett, P. R. (2009, September). Further moderators of the interrater reliability of supervisory ratings of job performance. Personnel Testing Council/Metropolitan Washington (PTC/MW) Newsletter, PTC/MW.

***Presentations***

Schwall, A. R., Beatty, A. S., & Jones, J. A. (2015, April). *Getting started with R: An interactive and hands-on tutorial.* Master tutorial presented at the annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Beatty, A. S., Buckley, K., Sprenger, A. M., & Russell, T. L. (2015, April). Using Mturk to pilot critical thinking items and guide test development. In S. D. Waters (Chair), *Advancing test development practices: Modern issues and technological advancements.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Fleischer, M. S., Putka, D. J., Beatty, A. S., McCloy, R. A., Woehr, D. J., & Rochette, L. M. (2015, April). Beyond personality level: Isolating trait expression in a single administration. In M. S. Fleischer (Chair), *Within-person variability of personality traits: Theory and applied measurement.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Schwall, A. R., Lustenberger, D. E., Beatty, A. S., & Jones, J. A. (2014, May). *Getting started with R: Examples and lessons learned.* Panel discussion at the annual conference of the Society of Industrial and Organizational Psychology, Honolulu, HI.

Koch, A. J., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Walmsley, P. T. (2014, May). Persistence in STEM among undergraduates: Why do women leave? In T. S. Behrend & D. A. Major (Chairs), *I-O psychology’s contributions to patching the STEM career pipeline.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Huber, C. R., Kuncel, N. R., Sackett, P. R., & Beatty, A. S. (2014, May). *Validity generalization across college applicant pools: A longitudinal study*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Barnett, R. C., & Beatty, A. S. (2013, April). The validity of assessor judgment in individual psychological assessment. In S. B. Morris & R. C. Barnett (Chairs), *Individual psychological assessment: Integrating science and practice.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Beatty, A. S., Berry, C. M., Barratt, C. L., & Sackett, P. R. (2013, April). *Testing the generalizability of indirect range restriction corrections.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2013, April). *Within-high-school vs. across-high-school predictors in admissions: Validity and diversity effects*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Beck, J. W., Beatty, A. S., & Sackett, P. R. (2013, April). *Normality of job performance: A reply to O’Boyle and Aguinis.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX. ***Featured Top Poster.***

Koch, A. J., Kuncel, N. R., Sackett, P. R., Beatty, A. S., & Walmsley, P. T. (2013, April). *Predicting persistence in science, technology, engineering, and math fields*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Koch, A. J., Sackett, P. R., Kuncel, N. R., Beatty, A. S., & Walmsley, P. T. (2013, April). *Exploring omitted variables: Why cognitive tests overpredict minority performance*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Walmsley, P. T., Kuncel, N. R., Sackett, P. R., Koch, A. J., & Beatty, A. S. (2013, April). *A method for displaying incremental validity with expectancy charts*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Beatty, A. S., Sackett, P. R., Kuncel, N. R., Kiger, T. B., Rigdon, J., & Shen, W. (2012, April). *An illustration of pareto-optimality applied to primary study educational data*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Beatty, A. S.,Trippe, D. M., & Russell, T. (2012, April). *A hybrid approach to developing parallel test forms*. In C. Paullin (Chair), *Practical IRT: Applications in real-world situations*. Symposium presented at the annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Shen, W., Kiger, T. B., Sackett, P. R., Kuncel, N. R., Walmsley, P., Beatty, A. S.,& Rigdon, J. L. (2012, April). *Joint effects of race and language in explaining differential prediction.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Beatty, A. S., Sackett, P. R., Kuncel, N. R., Kiger, T. B., Shen, W., & Rigdon, J. L. (2011, April). *Estimating the reliability of college grades.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Beatty, A. S., Sackett, P. R., Kuncel, N. R., Rigdon, J. L., Shen, W., & Kiger, T. B. (2011, April). *A comparison of two methods for keying biodata inventories*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Kiger, T. B., Kuncel, N. R., Sackett, P. R., Shen, W., Beatty, A. S., & Rigdon, J. L. (2011, April). *Exploring non-linearity in the relationship between HSGPA and college grades.* Poster presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.

Kiger, T. B., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Shen, W., & Rigdon, J. L. (2011, April). *A large sample response addressing low power in differential prediction.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Rigdon, J. L., Shen, W., Kuncel, N. R., Sackett, P. R., Beatty, A. S., & Kiger, T. B. (2011, April). *SES role in SAT-grade relationships across gender and racial subgroups.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2010, April). *“How am I doing?”: A meta-analytic review of 25 years of feedback-seeking research*. Poster presented at the annual conference of theSociety for Industrial and Organizational Psychology, Atlanta,GA.

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. (2010, June). *The nature and dimensionality of feedback-seeking behavior: A meta-analysis*. Poster presented at the 2nd Bi-annual IWP Conference on Work, Wellbeing and Performance, Sheffield, UK.

Rigdon, J., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Kiger, T. B., & Shen, W. (2010, April). *Considering SES in the use of standardized tests for selection*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA. ***Featured*** ***Top Poster***.

Shen, W., Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., & Kiger, T. B. (2010, April). *Moving beyond validity generalization: What about the remaining variance?* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Beatty, A. S., & Borneman, M. J. (2009, April). *Error rate of trim-and-fill methods to detect meta-analytic publication bias*. In M. J. Borneman (Chair), *Measurement concerns for the I/O practitioner (and researcher).* Symposium presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Beatty, A. S., & Sackett, P. R. (2009, April). *Revisiting differential prediction by gender for admissions testing.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Christensen, F. G. W., Seltzer, B. K., Beatty, A. S., & Ones, D. S. (2009, April). *Thirty years of meta-analysis: Assessing its impact on the sciences*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Shen, W., Beatty, A. S., & Sackett, P. R. (2009, April). *Further moderators of the interrater reliability of supervisory ratings of job performance.* Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA. ***Featured Top Poster***.

Beatty, A. S., Syzdek, M. R., Kellom, G., & Far, J. (2005, April). *The St. John’s experience: Misperception of other college men’s gender role conflict*. Poster presented at the 2005 convention of the American Men’s StudiesAssociation.

Immelman, A., & Beatty, A. S. (2005, July). *The political personality of 2004 U.S. presidential candidate John Kerry.* Poster presented at the 28th annual scientific meeting of the International Society ofPolitical Psychology, Toronto, Ontario.

Syzdek, M. R., Beatty, A. S., Kellom, G., & Far, J. (2005, August). *Misperceiving college men’s gender role conflict.* Poster presented at the 2005 convention of the American Psychological Association, Honolulu, HI.

Immelman, A., & Beatty, A. S.(2003, July). *The political personality of Zimbabwean president Robert Mugabe*. Poster presented at the 26th annual scientific meeting of the International Society of Political Psychology, Boston, MA.

Jeff Jones

Email: Jeff.Jones@KornFerry.com

**Education:**

Doctorate in Quantitative Psychology and Psychometrics, University of Minnesota, October, 2013. Advisor: Dr. Niels Waller.

Bachelor of Science, Psychology, University of California, Davis, June 2006.

Bachelor of Arts, Japanese Language and Literature, University of California, Davis, June 2006.

Minor in Mathematics, University of California, Davis, June 2006.

**Employment History:**

Senior Manager of Analytics, Korn Ferry, 2015 – Present.

Manager of Research and Analytics, Korn Ferry, 2013 – 2015.

Adjunct Professor, Hamline University, Fall 2013.

Graduate Instructor/Section Leader, University of Minnesota, 2006 – 2013.

**Awards:**

Korn Ferry Founder’s Award for Innovation, 2015.

Eva O. Miller Fellowship, 2012.

Graduate Summer Research Fellowship, 2009.

Graduate Research Partnership Program Fellowship, 2007.

**Publications:**

Jones, J. A. & Waller, N. G. (2016). Fungible weights in logistic regression.

*Psychological Methods, 21,* 241-260*.*

Jones, J. A. & Waller, N. G. (2015). The normal-theory and asymptotic distribution-free covariance

matrix of standardized regression coefficients: Theoretical extensions and finite sample behavior. *Psychometrika, 80,* 365-378.

Jones, J. A. & Waller, N. G. (2013). A comparison of methods for computing confidence intervals for

standardized regression coefficients. *Psychological Methods, 18,* 435-453.

Jones, J. A. & Waller, N. G. (2013). Abstract: The normal-theory and asymptotic distribution-free

covariance matrix of standardized regression coefficients. *Multivariate Behavioral Research, 48,*

161.

Waller, N. G. & Jones, J. A. (2011). Investigating the performance of alternate regression weights by

studying all possible criteria in regression models with a fixed set of predictors. *Psychometrika,*

*76,* 410-439.

Jones, J. A. (2010). GenCorr: An R routine to generate correlation matrices from a user-defined

eigenvalue structure. *Applied Psychological Measurement, 34,* 68-69.

Waller, N. G. & Jones, J. A. (2010). Correlation weights in multiple regression.  *Psychometrika, 75,*

58-69.

Waller, N. G. & Jones, J. A. (2009). Locating the extrema of fungible regression

weights. *Psychometrika, 74,* 589-602.

**Software:**

Goebl, A. P., Jones, J. A., Dahlke, J., & Beatty, A. S. (2016). iopsych: Methods for

industrial/organizational psychology. R package version 0.90.

Waller, N. G. & Jones, J. A. (2015). fungible: Fungible coefficients and Monte Carlo

functions. R package version 1.3.

**Presentations and Workshops:**

Jones, J. A., Goebl, A. P., & Semmel, S. G. (2016, April). *Handling big(gish) data in R:*

*An introductory and interactive tutorial.* Master Tutorial at the annual meeting of the Society of

Industrial and Organizational Psychology, Anaheim, CA.

Goebl, A. P. & Jones, J. A. (2016, April). *An R package for I-O psychology simulation*

*building: iopsych.* Poster presented at the annual meeting of the Society of

Industrial and Organizational Psychology, Anaheim, CA.

Goebl, A. P. & Jones, J. A. (2016, April). *Creative performance is a viable criterion for*

*personnel selection.* Poster presented at the annual meeting of the Society of

Industrial and Organizational Psychology, Anaheim, CA.

Blazek, S. & Jones, J. A. (2016, April). *A year on the job: Simulations-based*

*assessments’ versatile utility.* In D. Guangrong (Chair), *The art and science of executive assessment: Research and practice.* Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Anaheim, CA.

Blazek, S. & Jones, J. A. (2016, February). Streamlining your voice of the customer

program: Automating survey follow-ups, panel updates, and reports. Presentation given at Qualtrics Insight Summit, Salt Lake City, UT.

Jones, J. A. & Beatty, A. (2015, December). *R Workshop*. Workshop given at the

monthly meeting of Minnesota Professionals for Psychology Applied to Work, Minneapolis, MN.

Schwall, A., Beatty, A., & Jones, J. A. (2015, May). *Getting started with R: An*

*Interactive and Hands on Tutorial.* Master Tutorial at the annual meeting of the

Society of Industrial and Organizational Psychology, Philadelphia, PA.

D’Mello, S. & Jones J. A. (2015, May). *Demographic Differences in the Motivational*

*Drivers of Leaders*. Poster presented at the annual meeting of the Society of

Industrial and Organizational Psychology, Philadelphia, PA.

Goebl, A., Jones J. A., & Sharpe, E. (2015, May). *Relative Criteria Importance Depends*

*on Predictor Choices*. Poster presented at the annual meeting of the Society of

Industrial and Organizational Psychology, Philadelphia, PA.

Schwall, A., Lustenberger, D., Beatty, A., & Jones, J. A. (2014, May). *Getting started*

*with R: Examples and Lessons Learned.* Panel discussion at the annual meeting

of the Society of Industrial and Organizational Psychology, Honolulu, HI.

Huber, C. & Jones, J. A. (2014, May). *Principal components and factor scores in*

*multiple regression: A simulation.* Poster presented at the annual meeting of the

Society of Industrial and Organizational Psychology, Honolulu, HI.

Jones, J. A. & Waller, N. G. (2012). *The normal-theory and asymptotic distribution-free covariance*

*matrix of standardized regression coefficients.* Annual Meeting, Society for Multivariate Experimental Psychology Conference, Vancouver, British Columbia.

Morris P. E. & Jones, J. A. (2005). *Using formal inference-based recursive*

*modeling to detect plausible interactions for multiple regression.* Annual Meeting, American Psychological Society Conference, Los Angeles, CA.

**Technical Reports and White Papers:**

Blazek, E. S., Jones, J. A., Lewis, J. L, & Orr, J. E. (2016). Predicting financial gains.

Korn Ferry Institute, Korn Ferry.

Blazek, E. S., Jones, J. A., Lewis, J. L, & Orr, J. E. (2016). Leading indicators. Korn

Ferry Institute, Korn Ferry.

Blazek, E. S., Jones, J. A., Lewis, J. L, & Orr, J. E. (2016). Korn Ferry simulations-based

assessments predict CEO success: CEO outcomes research technical paper. Korn

Ferry Institute, Korn Ferry.

Dai, G., Davies, S., Goff, M., Jones J. A., D’Mello, S., Orr, J. E., Storfer, P., & Tang, K.

Y. (2014). Korn Ferry Leadership Architect: Research guide and technical manual. Version 14.2a—01/2016.<http://www.kornferry.com/technical-manuals>

Lewis, J. & Jones, J. A. (2016). Fit matters. Korn Ferry Institute, Korn Ferry.

<http://www.kornferry.com/institute/fit-matters>

Lewis, J., Goff, M., Jones, J. A., Hezlett S., Tang, K. Y., Dai, G., D’mello, S., Henry, L.,

Zes, D., Fetzer, J., Xie, C., & Scheer, P. (2015). Korn Ferry four dimensional executive

assessment: Research guide and technical manual. Version 15.1a—01/2016, Korn Ferry.<http://www.kornferry.com/technical-manuals>

Jones. J. A. & Waller, N. G. (2013). The normal-theory and asymptotic distribution-free

(ADF) covariance matrix of standardized regression coefficients: Theoretical

extensions and finite sample behavior. Technical Report 052513. University of

Minnesota, Twin Cities.

http://www.psych.umn.edu/faculty/waller/downloads/techreports/TR052913.pdf

**Manuscripts in Preparation:**

Davis-Stober, C. P., Waller, N. G., & Jones, J. A. (in preparation). A note on the accuracy of fixed

weighting schemes and fungible weights. (working title).

Jones, J. A., Waller, N. G., & Nydick, S. W. (in preparation). The cross-validation performance of

correlation weights in multiple regression.

Jones, J. A. & Waller, N. G. (in preparation). The normal-theory covariance matrix of standardized

regression coefficients derived from correlations and covariances. (working title).

Waller, N. G. & Jones, J. A. (in preparation). Confidence intervals for enhancement/suppression effects in

regression models with fixed and random predictors. (working title).

**Courses Taught:**

Introduction to Statistics

**Alexander R. Schwall, Ph.D.**

1343 Old Meadow Road, Pittsburgh, PA 15241

412-254-3030 alexander.schwall@gmail.com

**Education**

|  |  |  |
| --- | --- | --- |
| Ph.D. | Industrial/Organizational Psychology; The Pennsylvania State University | 2003-2008 |
| Diplom (Master’s) | Rheinisch-Westfälische Technische Universität Aachen, Germany  Graduated with distinction, Springorum Medal (highest level of distinction) | 2001-2003 |

**Applied Experience**

|  |  |
| --- | --- |
| Development Dimensions International (DDI), Pittsburgh, PA | **R&D Product Design Owner 3/2013 to present:**  Leading the development of a global, multi-purpose personality inventory, a battery-style leadership inventory for executive level candidates, and a leadership potential measure  Developing a novel aggregation method to combine assessment results across multiple tests and assessments (i.e. simulation) in R  Leading the maintenance of IRT based adaptive cognitive ability test, including item development and item calibration  Developing game-like, interactive test content  Serving as a technical contact for IT/Engineering to define technical specs for new interactive testing platform  Acting as internal/external consultant for legal and psychometric questions |
| Development Dimensions International (DDI), Pittsburgh, PA | **Delivery Consultant 10/2008 - 3/2013:**  Managed large scale job analyses and test implementation projects for global and multi-site clients  Developed customized multi-level and multi-language competency models for global clients  Managed multi-language test development and implementation projects with Fortune 500 companies  Consulted clients on development of recruiting and selection procedures, client specific company-wide competency models, and test customization |
| Vance & Renz,  State College, PA | **Researcher 05/2007 – 09/2007:** Co-developed and executed large scale (n>100k) longitudinal analysis of traffic safety data |
| SHL GmbH,  Hamburg, Germany | **Freelancer 05/2001-06/2003:** Administered personnel selection tests and AC exercises in Germany and UK to assessment center candidates |
| Roland Berger Strategy  Consultants,  Munich, Germany | **Internship 07/2002-10/2002:** Independently statistically analyzed multinational employee survey and participated in the evaluation and improvement of an internet commerce platform for a German Bank |

**Teaching & Research**

|  |  |
| --- | --- |
| Research Assistant, State College, PA | **08/2007—08/2008:** Coordinated, administered and analyzed qualitative interviews, participated in the development of theoretical model of creativity and innovation |
| Teaching fellow, State College, PA | **08/2006-05/2007, Spring 2005, 2006:** Instructed 400 level class on Personnel Selection, ran statistics lab for graduate students in psychology department |

**Graduate Coursework**

|  |  |  |
| --- | --- | --- |
| Personnel Testing and Selection  Performance Appraisal  Micro Organizational Behavior  Macro Organizational Behavior  Training and Development | Data Analysis I  Data Analysis II  Decision Making in Groups  Innovation and Creativity | Current Topics in Social Psychology  Personnel Development  Regression  Multivariate Change  Multilevel Modeling |

**Selected Publications & Presentations**

Schwall, A. R. (2015). The Triangle Item: A Graphical User Interface for Frequency Items. In M. S. Fleisher (Chair), Within-Person Variability of Personality Traits: Theory and Applied Measurement. Symposium at the 30th annual meeting of the Society for Industrial-Organizational Psychology, Philadelphia, PA.

Schwall, A., Beatty, A., & Jones, J. A. (2015, May). *Getting started with R: An Interactive and Hands on Tutorial*. Master Tutorial at the annual meeting of the Society of Industrial and Organizational Psychology, Philadelphia, PA.

Schwall, A. R., Lustenberger, D., Beatty, A., & Jones, J. A. (2014, May). *Getting started with R: Examples and Lessons Learned.* Panel discussion at the annual meeting of the Society of Industrial and Organizational Psychology, Honolulu, HI.

Schwall, A. R. (2014). Preventing Faking in Biodata Items using Elaboration and Machine Learning. In C. M. Reddock & A. R. Schwall (Co-chairs), New Developments in Biodata Research and Practice. Symposium at the 29th annual meeting of the Society for Industrial-Organizational Psychology, Honolulu, Hawaii.

Schwall, A. R. (2014). IGNITE and Discuss: Perspectives on High Potential (IGNITE Discussion with Adler, S., Conte, J. M., Gibbons, A. M., Ginther, N. G., Jacobs, R., Kinney, T., Lemelle, C., Silzer, R., Warrenfeltz, R.). Presented at the 29th annual meeting of the Society for Industrial-Organizational Psychology, Honolulu, Hawaii.

Schwall, A. R. (2012). Defining Age, and using Age-relevant Constructs. In J. W. Hedge & W. C. Borman (Eds.), *The Oxford Handbook of Work and Aging.* (pp.169-186).Oxford:Oxford University Press

Schwall, A. R., Giumetti, G. W., Schmidt, D. B., Sinar, E. F. (2010, April). Adverse impact in large samples: Differing conclusions depending on methodology. In E. F. Sinar & J. Morrison (Co-chairs), The 4/5ths is Just a Fraction: Alternative Adverse Impact Methodologies. Symposium presented at the 25th annual meeting of the Society for Industrial-Organizational Psychology, Atlanta, GA.

Mohammed, S., & Schwall, A. R. (2009). Individual differences and decision making: What we know and where we go from here. In G. P. Hodgkinson & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology, 24* (pp. 249-312). John Wiley & Sons

Farr, J. L. & Schwall, A. R. (2008). New employment structures: Varieties of impact on aging workers. In Schaie, W. & Abeles, R. (Eds.), *Social Structures and Aging Individuals: Continuing Challenges*, pp. 213-230. New York: Springer.

1. Note that at the time of submission the associated file location only contains sample materials to demonstrate the general process. [↑](#footnote-ref-1)