



Gender differences in work sample assessments: Not all tests are created equal

Michael B. Harari^{a*}, Chockalingam Viswesvaran^a, and Rachel O'Rourke^b

^aFlorida International University, USA

^bCity of Jacksonville, USA

ARTICLE INFORMATION

Manuscript received: 10/10/2013

Revision received: 21/02/2014

Accepted: 06/03/2014

Keywords:

Work samples

Gender differences

Typing skills

Verbal comprehension

Palabras clave:

Muestras de trabajo

Diferencias de género

Destrezas mecanográficas

Comprensión verbal

ABSTRACT

We examined gender differences in five different types of clerical work sample exercises. Our results indicated that, overall, gender differences were minimal. However, there was substantial variability in gender differences across work sample exercises such that the magnitude and direction of gender differences depended on the constructs that each assessed. Gender differences favored females on tests that involved typing skills and verbal comprehension, while gender differences favored males on tests that involved spatial abilities. We found mixed support for the notion that males would outperform females on work sample assessments of technical skills.

© 2014 Colegio Oficial de Psicólogos de Madrid. Production by Elsevier España, S.L. All rights reserved.

Diferencias de género en la evaluación de muestras de trabajo: no todas las pruebas se crean por igual

RESUMEN

Analizamos las diferencias de género en cinco tipos distintos de ejercicios de muestras de trabajo administrativo. Los resultados indican que, en general, las diferencias de género eran mínimas. No obstante, había gran variabilidad en las diferencias de género en los ejercicios de muestras de trabajo en función de los constructos evaluados por cada uno de ellos. Las diferencias de género favorecían a las mujeres en las pruebas que implicaban destrezas mecanográficas y comprensión verbal, mientras que las diferencias de género favorecían a los hombres en pruebas que implicaban aptitudes espaciales. No hubo un apoyo claro con respecto a la noción de que los hombres superan a las mujeres en la evaluación de muestras de trabajo de capacidades técnicas.

© 2014 Colegio Oficial de Psicólogos de Madrid. Producido por Elsevier España, S.L. Todos los derechos reservados.

Work samples are assessments that simulate tasks and cognitive processes performed on the job (Ployhart, Schneider, & Schmitt, 2006). That is, rather than being designed to assess a single underlying construct (such as assessments of cognitive ability or personality), they are designed to have high content validity, assessing proficiencies in tasks required for successful performance on the job, and therefore assess a range of constructs (Roth, Bobko, McFarland, & Buster, 2008). Meta-analytic reviews have indicated that work samples have high criterion-related validity for predicting job performance (Roth, Bobko, & McFarland, 2005; Schmidt & Hunter, 1998). Furthermore, as work samples have a higher degree of fidelity to tasks performed on the job than traditional individual differences predictors (Asher & Sciarino, 1974), they are generally associated

with more favorable applicant reactions (e.g., Hausknecht, Day, & Thomas, 2004; Macan, Avedon, Paese, & Smith, 1994).

While criterion-related validity and applicant reactions are important considerations for predictors used in personnel selection, another important criterion is mean group differences (Pyburn, Ployhart, & Kravitz, 2008). For example, cognitive ability is among the most robust predictors of job performance across occupations and criteria (cf. Schmidt & Hunter, 1998); however, there are subgroup differences in cognitive ability scores such that minority group members generally score lower than majority group members (Bobko & Roth, 2013; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Indeed, a number of predictors traditionally used in personnel selection show group differences that result in adverse impact for minority members of various racial, ethnic, and gender subgroups (Ployhart & Holtz, 2008).

A purported benefit of work sample assessments is a reduction in group differences as compared to more traditional assessments, such as cognitive ability tests (Schmidt, Greenhal, Hunter, Berner, &

*Correspondence concerning this article should be addressed to Michael B. Harari. Department of Psychology. Florida International University. Miami, FL. USA 33199. E-mail: mharari@fiu.edu

Seaton, 1977; Schmitt, Clause, & Pulakos, 1996). However, recent research suggests that prior estimates of subgroup differences in work sample scores may be understated (Bobko, Roth, & Buster, 2005). Specifically, the majority of research in this area was done using samples of job incumbents rather than applicants and as a result, estimates of subgroup differences have been deflated due to range restriction. For instance, once accounting for the effect of range restriction, research has indicated that black-white mean differences in work sample assessments may be equal to that of cognitive ability tests (Roth et al., 2008).

While revised estimates (i.e., accounting for range restriction) of ethnic group differences in work sample scores appear in the literature (Bobko & Roth, 2013; Roth et al., 2008), much less research has considered this issue for gender differences. Indeed, the few studies that have investigated gender differences in work samples (e.g., Hatstrup & Schmitt, 1990; Pulakos, Schmitt, & Chan, 1996) have been conducted using samples of incumbents and as a result, their findings are biased due to range restriction (Bobko et al., 2005). Additionally, while work samples assess a heterogeneous set of constructs, research has rarely considered that gender differences might vary between assessments based on the constructs that they assess. A rare exception is a study conducted by Roth, Buster, and Barnes-Farrell (2010) that reported subgroup differences in work sample tests using a sample of job applicants. This research considered gender differences in work sample tests that assessed the following skills: technical, social, and written. The results of this study indicated that, overall, gender differences in work sample scores favored females. However, there was substantial heterogeneity across exercises such that the direction and magnitude of gender differences depended on the constructs underlying performance on each assessment. Specifically, males outperformed females on tests that assessed technical skills, while females outperformed males on tests that assessed social and written skills.

The purpose of the present study is to extend this line of inquiry by examining gender differences in work sample assessments used to predict performance in clerical positions, including assessments of technical skills, typing skills, verbal comprehension, and spatial abilities. While assessments of these constructs are frequently used in applied contexts for selecting personnel into clerical positions (Pearlman, Schmidt, & Hunter, 1980; Whetzel et al., 2011), with the exception of technical skills (Roth et al., 2010), research has not considered gender differences in work sample assessments of these constructs. We address this gap in the present study. Consistent with prior research (Roth et al., 2010), we propose that gender-based subgroup differences will vary across exercises depending on the constructs that they assess. In the following sections, we discuss the constructs involved in the clerical work samples examined in the present study, review the literature concerning gender differences in these constructs, and make predictions concerning gender differences in assessment scores.

Constructs Involved in Work Sample Tests

A wide number of constructs are likely to be assessed by work samples and the construct validity of these assessments may vary greatly across exercises (Roth et al., 2008). Our review will focus specifically on individual differences in skills and abilities that are involved in the clerical assessments considered in the present study that are relevant when considering gender subgroup differences – technical skills, typing skills, verbal comprehension, and spatial abilities. As prior research in this area has already considered technical skills (Roth et al., 2010), we begin our review with this construct. We then review skills and abilities that, while commonly used to select clerical personnel in applied settings (Pearlman et al., 1980; Whetzel et al., 2011), have not been the focus of research into mean gender differences (i.e., typing skills, verbal comprehension, spatial abilities).

Technical skills. Technical skills refer to task-related competencies. Roth et al. (2010) drew on a number of theories to propose that males would, on average, score higher than females on work sample assessments of technical skills. For instance, males are more likely to develop independent self-construals than women (Cross & Madson, 1997a, 1997b), which facilitate a focus on developing technical competencies in order to distinguish oneself from others. Furthermore, men are more likely to possess *agentic orientations* (Eagly & Johannesen-Schmidt, 2001), which results in a tendency to learn about the world through task-oriented behaviors. Consistent with these theoretical perspectives and prior research, we propose that males will score higher than females on exercises that involve technical skills.

Hypothesis 1: Males will score higher than females on work sample assessments that involve technical skills.

Typing skills. Typing skills refer to the speed and accuracy with which one transcribes or composes text using a keyboard (Salthouse, 1985). For the purposes of the present study, we consider transcription typing only (i.e., reproducing a typed copy of a stimulus), as opposed to composition typing (i.e., typing original work). A number of theoretical models have been developed to explicate the various cognitive and motor processes involved in transcription typing (Salthouse, 1986). Specifically, research and theory related to transcription typing suggest that the cognitive processes involved make use of verbal working memory (Hayes & Chenoweth, 2003; Kellog, 1999; Levy & Marek, 1999), and research suggests that females perform superior to males on verbal working memory tasks (Speck et al., 2000). Transcription typing also requires the implementation of precise motor movements with speed and accuracy (Salthouse, 1986), and research suggests that females outperform males on such fine motor coordination tasks (Thomas & French, 1985). As a result, we propose that females will score higher than males on exercises that involve typing skills.

Hypothesis 2: Females will score higher than males on work sample assessments that involve typing skills.

Verbal comprehension. Consistent with the demands of the clerical job considered here, several of the work sample assessments required the applicant to listen to prerecorded information and either (a) dictate what they heard or (b) recall the information at a later point to answer questions. Such tasks involve verbal working memory and fine motor coordination, as already discussed. However, both also involve listening to, comprehending, and, in some cases, remembering and recalling auditory stimuli. Research suggests that females tend to outperform males in these types of verbal comprehension tasks (Herlitz, Nilsson, & Backman, 1997; Hough, Oswald, & Ployhart, 2001). For instance, females are superior to males in terms of word recognition (Temple & Cornish, 1993) and recalling details from short stories (Hultsch, Masson, & Small, 1991; Zelinski, Gilewski, & Schaie, 1993). Therefore, we propose that females will score higher than males on work samples that involve verbal comprehension.

Hypothesis 3: Females will score higher than males on work sample assessments that involve verbal comprehension.

Spatial abilities. Spatial abilities include the ability to determine spatial relationships and to mentally rotate or manipulate two- and three-dimensional figures (Linn & Peterson, 1985). The results of a number of quantitative and qualitative reviews converge in suggesting that males outperform females on these types of tasks (Hough et al., 2001; Maccoby & Jacklin, 1974; Voyer, Voyer, & Bryden, 1995). Spatial abilities (specifically related to determining spatial

relationships) may be invoked in clerical work sample assessments that, for example, ask applicants to read and interpret maps to respond to caller requests.

Hypothesis 4: Males will score higher than females on work sample assessments that involve spatial abilities.

In summary, the purpose of the present study is to assess gender differences in work sample assessments for clerical jobs. While work samples that assess constructs such as typing skills, verbal comprehension, and spatial abilities have been used to select personnel into clerical occupations for some time (Pearlman et al., 1980), we are aware of no research that has assessed gender differences in these assessments. Thus, the gender adverse impact potential associated with some common work sample predictors used in a clerical context are presently unknown and we address this issue in the present study.

Method

Database

The database used to conduct this study was provided by a government agency in the southeastern United States. The database contained scores on five work sample exercises (discussed below) for applications to an emergency call center operator position. The testing was done during two months: March 2011 and July 2011. For purposes of the present study, we pooled these databases together. In March, 88 applicants were tested, while in July, 176 applicants were tested, resulting in a total of $N = 264$. The testing procedures differed slightly between March and July. In March, a compensatory method was used whereby each applicant was administered all predictors and thus, no direct range restriction was involved. On the other hand, in July, participants were first prescreened using the typing test (discussed below) before taking the remainder of the predictor battery, and thus, there was direct range restriction on this predictor. Specifically, of the 176 applicants tested, 64 were eliminated due to their scores on the typing test.

The presence of direct range restriction was troublesome, as it may have biased our estimates of subgroup differences (Bobko et al., 2005). However, we had the necessary data available to account for the influence of range restriction on our findings. When dividing the standard deviation of typing scores in the entire applicant sample by the standard deviation of typing scores in the subsample that took the entire selection battery, the resulting $U_x = 1.05$.

Gender differences were assessed using the sample of applicants that were administered all predictors ($n = 200$). However, analyses involving the typing assessment were based off of the entire sample of $N = 264$. The sample size and gender composition of the entire applicant pool, the applicants who were screened out based on their typing scores, and the applicants who were administered the entire battery, is indicated in Table 1.

Table 1

Sample size and gender composition of applicant pool

	Entire applicant pool	Screened out	Administered all predictors
N_{Total}	264	64	200
N_{Male}	57 (21.6%)	17 (26.6%)	40 (20.0%)
N_{Female}	207 (78.4%)	47 (73.4%)	160 (80.0%)

Measures

Each applicant was administered a selection battery comprised of five work sample exercises. Each exercise was selected based on its

relevance to key competencies required for successful job performance. Four of the five assessments were licensed through a testing vendor and one was developed internally by the organization. All were proprietary and thus we were provided only scale scores rather than item level data and were unable to compute internal consistency reliability estimates. All exercises were administered using a computer and were objectively scored.

Technical skills. This work sample assessed technical performance in clerical positions. Specifically, applicants read descriptions of problems that were typical of those encountered in clerical positions. For instance, the applicant faced issues related to shipping a package, filing, prioritizing phone calls, and scheduling a meeting. Applicants had to determine a solution to the problem using the information available to them and indicate a response to the scenario in multiple-choice format. As this assessment was multiple-choice, applicants were assessed based on their resolution to the problem, rather than the behaviors that they had engaged in in order to do so.

Typing skills. Typing skills were assessed in two work sample tests: a typing test and a data entry test.

Typing. A standard typing test was employed by the organization. Applicants were presented with two passages that they were instructed to transcribe onto a computer. Scores were calculated based on speed and accuracy and expressed as words-per-minute.

Data entry. A data entry work sample assessment was employed by the organization. Applicants wore headphones that presented them with information related to a simulated caller's name, job title, and other personal information. Applicants were instructed to type this information into a computer database. Scores were calculated based on speed and accuracy and expressed as words-per-hour.

Verbal comprehension. The verbal comprehension assessment included two components. The first was an assessment of listening skills. The applicants first listened to an audio clip that explained a situation and were then asked questions about it. This assessment was administered in multiple-choice format. The second component required the applicant to gather information by reading the computer screen and listening to a simulated caller. The applicants were to use this information to answer multiple-choice questions and to fill in data fields in a computer database. The applicants faced realistic issues in this task, such as a poor phone connection and distressed callers. While this assessment is composed of two exercises, only composite scores were available.

Spatial ability. This work sample involved reading maps and was developed by the organization in order to assess spatial reasoning and knowledge of local roadways. Applicants were provided with a paper copy of a map, which they were to read and interpret in order to answer a series of multiple-choice questions pertaining to, for example, plotting routes and identifying intersections.

Analyses

To assess the presence of subgroup differences in work sample scores, we computed d values, an index of standardized mean differences in scores between two groups. Specifically, we calculated d values by subtracting the mean assessment score for females from the mean assessment score for males, and divided this value by the standard deviation pooled across groups. Thus, positive d values indicated subgroup differences favoring males, while negative d values indicated subgroup differences favoring females. To assess hypotheses 1 through 4, we calculated the d value associated with each of the five exercises. In interpreting the magnitude of d values,

we refer to Cohen's (1992) rules of thumb – values of 0.20, 0.50, and 0.80, indicate small, medium, and large standardized mean differences, respectively. We also assessed subgroup differences in unit-weighted composite scores across the five exercises. We formed composites by averaging standardized scores across each exercise for the $n = 200$ applicants who were administered all assessments. As there was direct selection on the typing skills assessment, we present both observed d values (i.e., D) and corrected d values (i.e., d). Corrections for direct range restriction were made using the formulas provided by Bobko, Roth, and Bobko (2001). We did not report truncated estimate of gender differences in typing scores because this value was estimated using the entire applicant pool.

Results

Means, standard deviations, intercorrelations among study variables, and d values are reported in Table 2. As the differences between corrected (i.e., d) and uncorrected (i.e., D) estimates of standardized gender differences are similar (i.e., they differ by only as much as 0.01), we review corrected estimates only. Hypothesis 1 predicted that gender differences would favor males in work samples that involved technical performance (i.e., the technical skills work sample). While males did outperform females on this assessment ($d = 0.11$), this difference was small in magnitude (Cohen, 1992) and was not statistically significant. Thus, hypothesis 1 was not supported.

Hypothesis 2 predicted that gender differences in work samples that involved typing skills would favor females. This hypothesis was supported, as females outperformed males on both the typing and data entry assessments ($d = -0.45$ and -0.42 for the typing and data entry test, respectively). Furthermore, the standardized mean differences were medium in magnitude (Cohen, 1992) and statistically significant. Hypothesis 3 predicted that females would outperform males on assessments that involved verbal comprehension. Our results provided mixed support for this prediction. Specifically, while the standardized mean differences on the verbal comprehension test favored females and was small-to-medium in magnitude ($d = -0.33$), it was not statistically significant. Finally, hypothesis 4 predicted that males would score higher than females on assessments involving spatial abilities. This hypothesis was supported as the standardized mean difference on the spatial ability test ($d = 0.42$) was medium in magnitude (Cohen, 1992), statistically significant, and favored males. Finally, in terms of the unit-weighted composite, the standardized mean difference (both uncorrected and corrected for range restriction) was $d = -0.22$. Thus, overall assessment scores favored females. However, this difference was not statistically significant.

Discussion

Research into adverse impact associated with predictor scores is of the utmost importance for personnel selection practices (Bobko & Roth, 2013). However, research in this area has faced a number of limitations. First, many studies have assessed group differences in predictor scores using incumbent samples and thus, prior estimates of such differences have been underestimated due to range restriction (Bobko et al., 2005). Second, little research has examined subgroup differences in work sample scores and of those that have, many have treated work samples as assessments of a homogenous construct, rather than attempting to assess between exercise differences in subgroup differences based on the constructs that they assess (Roth et al., 2008). Third, while some research has addressed these deficits in relation to ethnic subgroup differences in work sample scores (cf. Bobko & Roth, 2013), very little research has done so when considering gender-based group differences (Roth et al., 2010). Finally, no research has examined mean gender differences in work sample assessments of constructs that are frequently used to select personnel into clerical positions (Pearlman et al., 1980; Whetzel et al., 2011). In the present study, we addressed these gaps by examining gender differences in a diverse array of clerical work sample tests using a sample of job applicants.

The results of this study indicated that, across assessments, gender-based group differences in clerical work sample scores were minimal. Across exercises, we observed a standardized group difference of $d = -0.22$, and thus, on average, females scored slightly higher than males on these assessments. However, and consistent with our predictions, there was substantial variability in d values across exercises depending on the skills and abilities involved in performance on each assessment.

Counter to our predictions, we observed little evidence that males outperformed females on assessments of technical skills. While the mean differences in scores on the technical skills predictor did favor males ($d = 0.11$), the difference was small in magnitude and was not statistically significant. This is somewhat consistent with past research that also found only mixed support for proposed gender differences in assessments of technical skills (Roth et al., 2010). It is possible that technical skills are still too broad to reveal meaningful subgroup differences in work sample scores. For example, a number of distinct constructs likely influence technical skills, such as prior experience, knowledge, and cognitive abilities (e.g., Schmidt, Hunter, & Outerbridge, 1986), and it would perhaps be useful to consider the saturation of work samples with these constructs as opposed to technical skills per se. For example, the technical skills work sample employed in the present study

Table 2
Means, standard deviations, and intercorrelations among study variables

	Mean	SD	1	2	3	4	D	d
1 Technical skills	6.77 (6.98/6.72)	2.32 (2.33/2.32)					.11	.11
2 Typing ^a	38.30 (33.98/39.49)	13.06 (11.13/13.31)	.16*				--	-.45**
3 Data entry	4706.63 (4289.48/4810.92)	1331.17 (1197.56/1345.85)	.24**	.53**			-.41*	-.42*
4 Verbal comprehension	19.37 (18.70/19.53)	2.55 (2.76/2.48)	.41**	.26**	.34**		-.32	-.33
5 Spatial ability	7.13 (7.98/6.92)	2.72 (2.43/2.76)	.36**	.21**	.35**	.18**	.41*	.42*

Note. Means and standard deviations for males and females, respectively, appear in parentheses below the total sample estimates; ^abased off of $N = 264$.

may have been heavily saturated with cognitive ability, as this assessment required applicants to solve problems that involved the application of both verbal and quantitative abilities. Considering this, our finding of small gender differences is consistent with research indicating minimal gender differences in cognitive ability assessments (cf. Hough et al., 2001).

The other d values computed in the present study largely supported our predictions. First, females outperformed males in typing assessments. Second, while we did not identify statistically significant group differences in the verbal comprehension assessment, the d value was small-to-medium in magnitude (Cohen, 1992) and in the predicted direction. Finally, males outperformed females on assessments that involved spatial abilities. Consistent with other work in this area (Roth et al., 2010), our findings suggested that in order to understand gender differences in work sample assessments, careful consideration of the constructs involved in the work sample is required.

Limitations and Future Directions

A limitation of the present study is that many work samples assessed a number of constructs in addition to the predominant construct assessed. For instance, the verbal comprehension test involved typing as well. However, this limitation is inherent in all work samples. Nonetheless, future research into gender differences in work sample assessments that involve verbal comprehension should attempt to separate this construct from others that may be involved in work sample assessments that are relevant when considering gender differences in scores (e.g., typing skills).

Another limitation is that there was direct range restriction on typing scores for a subset of our sample. While we were able to account for this issue using range restriction corrections, the adequacy of these corrections depends on the assumptions of linearity and homoscedasticity being met in the data (Bobko et al., 2001). Therefore, while corrections are useful for estimated disattenuated effect sizes, future research is needed to assess gender differences in work sample assessments using samples of job applicants where no direct range restriction exists. However, we also note that the degree of range restriction in the present study was minimal ($U_x = 1.05$) and therefore the biasing effect on our estimates was not great.

Conclusion

Investigations into subgroup differences of predictors used in personnel selection are of critical importance for the scientific practice of I-O Psychology (Bobko & Roth, 2013). A gap in this literature concerns gender differences in work sample tests (Roth et al., 2010), and specifically those used for clerical positions. By using a sample of job applicants where minimal range restriction existed, we were able to more accurately assess the magnitude of gender-based subgroup differences in work sample scores than prior research. Furthermore, by taking a construct-oriented approach rather than treating work sample tests as a homogenous class of predictors (Roth et al., 2008) we were able to identify that the extent of gender differences in work sample scores depended on the constructs assessed by the respective assessment. Cumulatively, our findings suggested that gender differences in work sample assessments are not problematic so long as the assessments cover a wide-range of constructs. However, to ensure the minimization of subgroup differences in work sample scores, practitioners should carefully consider the constructs involved in each assessment.

Conflict of Interest

The authors of this article declare no conflicts of interest.

References

- Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, 27, 519–533. doi: 10.1111/j.1744-6570.1974.tb01173.x
- Bobko, P., & Roth, P. (2013). Reviewing, categorizing, and analyzing the literature on black-white mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology*, 66, 91–126. doi: 10.1111/peps.12007
- Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of d for range restriction and unreliability. *Organizational Research Methods*, 4(1), 46–61. doi: 10.1177/109442810141003
- Bobko, P., Roth, P. L., & Buster, M. A. (2005). Work sample selection tests and expected reduction in adverse impact: A cautionary note. *International Journal of Selection Assessment*, 13, 1–10. doi: 10.1111/j.0965-075X.2005.00295.x
- Chenoweth, A., & Hayes, J. (2003). The inner voice in writing. *Written Communication*, 20, 99–118.
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, 112(1), 155–159. doi: 10.1037//0033-2909.112.1.155
- Cross, S. E., & Madson, L. (1997a). Models of the self: Self-construals and gender. *Psychological Bulletin*, 112, 5–37. doi: 10.1037//0033-2909.112.1.5
- Cross, S. E., & Mason, L. (1997b). Exploration of models of the self: Reply to Baumeister & Sommer (1997) and Martin & Ruble (1997). *Psychological Bulletin*, 122, 51–55.
- Eagly, A. H., & Johannesen-Schmidt, M. C. (2001). The leadership styles of women and men. *Journal of Social Issues*, 57, 781–797. doi: 10.1111/0022-4537.00241
- Hatrup, K., & Schmitt, N. (1990). Prediction of trades apprentice's performance on job sample criteria. *Personnel Psychology*, 43, 453–466.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639–683. doi: 10.1111/j.1744-6570.2004.00003.x
- Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23, 135–149. doi: 10.1177/0741088306286283
- Herlitz, A., Nilsson, L., & Backman, L. (1997). Gender differences in episodic memory. *Memory and Cognition*, 25, 801–811. doi: 10.3758/BF03211324
- Hough, L., Oswald, F., & Ployhart, R. (2001). Adverse impact and group differences in constructs, assessment tools, and personnel selection procedures: Issues and lessons learned. *International Journal of Selection and Assessment*, 9, 152–194.
- Hultsch, D. F., Masson, M. E. J., & Small, B. J. (1991). Adult age differences in direct and indirect tests of memory. *Journal of Gerontology: Psychological Sciences*, 46, 22–30. doi: 10.1093/geronj/46.1.P22
- Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance & G. C. Jeffrey (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 42–61). Amsterdam, The Netherlands: Amsterdam University Press.
- Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of the phonological loop. In M. Torrance & G. C. Jeffrey (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 42–61). Amsterdam, The Netherlands: Amsterdam University Press.
- Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of gender differences in spatial abilities: A meta-analysis. *Child Development*, 56, 1479–1498.
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, 715–738. doi: 10.1111/j.1744-6570.1994.tb01573.x
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373–406. doi: 10.1037//0021-9010.65.4.373
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing race/ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172. doi: 10.1111/j.17446570.2008.00109.x
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pulakos, E. D., Schmitt, N., & Chan, D. (1996). Models of job performance ratings: An examination of rater race, rater gender, and rater level effects. *Human Performance*, 9, 103–119. doi: 10.1207/s15327043hup0902_1
- Pyburn, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61, 143–151. doi: 10.1111/j.1744-6570.2008.00108.x
- Roth, P. L., BeVier, C. A., Bobko, P., Switzer, F. S. III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330. doi: 10.1111/j.1744-6570.2001.tb00094.x
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, 58, 1009–1037. doi: 10.1111/j.1744-6570.2005.00714.x
- Roth, P., Bobko, P., McFarland, L., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology*, 61, 637–662. doi: 10.1111/j.1744-6570.2008.00125.x
- Roth, P. L., Buster, M. A., & Barnes-Ferrell, J. (2010). Work sample exams and gender adverse impact potential: The influence of self-concept, social skills, and written skills. *International Journal of Selection and Assessment*, 18(2), 117–130. doi: 10.1111/j.1468-2389.2010.00494.x
- Salthouse, T. A. (1985). Anticipatory processing in transcription typing. *Journal of Applied Psychology*, 70(2), 264–271. doi: 10.1037//0021-9010.70.2.264

- Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*, 99, 303-319. doi: 10.1037//0033-2909.99.3.303
- Schmidt, F. L., Greenthal, A. L., Hunter, J. E., Berner, J. G., & Seaton, F. W. (1977). Job samples vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. *Personnel Psychology*, 30, 187-197. doi: 10.1111/j.1744-6570.1977.tb02088.x
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274. doi: 10.1037//0033-2909.124.2.262
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432-439. doi: 10.1037//0021-9010.71.3.432
- Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. *International Review of Industrial and Organizational Psychology*, 11, 115-139.
- Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Change, L. (2000). Gender differences in the functional organization of the brain for working memory. *Neuroreport*, 11, 2581-2585. doi: 10.1097/00001756-200008030-00046
- Temple, C. M., & Cornish, K. M. (1993). Recognition memory for words and faces in schoolchildren: A female advantage for words. *British Journal of Developmental Psychology*, 11, 421-426. doi: 10.1111/j.2044-835X.1993.tb00613.x
- Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin*, 98, 260-282. doi: 10.1037//0033-2909.98.2.260
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250-270. doi: 10.1037//0033-2909.117.2.250
- Whetzel, D. L., McCloy, R. A., Hooper, A., Russell, T. L., Waters, S. D., Campbell, W. J., & Ramos, R. A. (2011). Meta-analysis of clerical performance predictors: Still stable after all these years. *International Journal of Selection and Assessment*, 19(1), 41-50. doi: 10.1111/j.1468-2389.2010.00533.x
- Zelinski, E. M., Gilewski, M. J., & Schaie, K. W. (1993). Individual differences in cross-sectional and 3-year longitudinal memory performance across the adult life span. *Psychology and Aging*, 8, 176-186. doi: 10.1037//0882-7974.8.2.176