UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Personnel selection

Mol, S.T.

DOI
10.4337/9781803921761.00102
Publication date
2024
Document Version
Final published version
Published in
Elgar Encyclopedia of Organizational Psychology

License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
Mol, S. T. (2024). Personnel selection. In P. M. Bal (Ed.), *Elgar Encyclopedia of Organizational Psychology* (pp. 518-522). (Elgar Encyclopedias in Business and Management). Edward Elgar Publishing. https://doi.org/10.4337/9781803921761.00102

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Nov 2026

Personnel selection

Description of the topic

An important prerequisite for personnel selection to be viable is the recruitment of a sufficient number of candidates to select from, something that may be getting to be less and less realistic for particular jobs in these times of (i) changing demographics associated with baby boomers across many industrialized nations reaching retirement age (implying, for instance, greater demands on care being placed on the shoulders of a shrinking workforce), (ii) increasing specialization (in part due to the automation of those work activities that are relatively structured and straightforward) and (iii) widespread political resistance against strategically facilitating labor migration. Stated in the extreme, if there is only one candidate for the position at hand, that person is going to have to be selected no matter what their (predicted) suitability may be.

Personnel selection can be defined as the art of ranking on a single dimension (from most to least suitable) a set of candidates on the basis of one or more relatively stable individual characteristics. These characteristics (such as personality, expertise or ability; henceforth individual differences) need to be determined a priori to be related to job performance, so that organizational decision makers can make evidence-based decisions as to who is most likely to exhibit the best job performance once hired. In essence then, the traditional conception of personnel selection comes down to the prediction of future (i.e. unknown) individual job performance on the basis of quantified (or quantifiable) individual differences that are – or become – known during the selection process, and that are combined into a single index on which all applicants can be ranked. Means of gauging such individual differences may include application materials, structured interviews and/or a formal assessment.

Definitions

Before pressing on with a more critical reflection on the art of personnel selection, it may be useful to provide a number of definitions beyond the definition of personnel selection that was provided at the start of this entry. Here, due to space limitations, I will restrict myself to discussing the terms *measurement error*, *construct validity*, *ecological validity*, *external validity*, *practical validity*, *false positives*, *false negatives*, *job description*, *sampling error*, *job performance*, *criterion measure*, *predictor measure*, *content validity*, and *criterion-related validity*.

The fact that personnel selection ultimately comes down to a probabilistic prediction problem means that even in the best of scenarios errors are going to be made. The prime reason for this is that in practice a perfect correlation (i.e. a straight-line relationship) between some individual-differences-based index and performance scores is an unattainable ideal. This means that any threshold on the index that is used to select individuals is going to be associated with imprecise estimates of what is essentially unknown performance at the time of selection. Reasons for such imprecision derive from (i) measurement error on the predictor and/or performance measures in the a priori validation effort (i.e. poor construct validity); (ii) non-interchangeability between the job or the context for which the selection measure was developed and validated, and the job or context in which it is being leveraged to make personnel selection decisions (poor ecological validity); (iii) non-interchangeability between the individuals on whom the selection measure was developed and validated and the individual(s) to whom it is applied (poor external validity); and (iv) managerial resistance to, and/or the practical impossibility of, investing in systematically researching, evidencing and validating a selection program for each distinct (and changeable) job in the organization (something we could refer to as poor practical validity).

In thinking about the types of errors that can be made in personnel selection (for a more comprehensive treatment the reader is referred to Roe, 2013), it is useful to distinguish between the theoretical notions of false positives (those individual candidates who surpass the threshold on the selection-context measures of individual differences, but fail to surpass some preset acceptability threshold on job performance once hired) and false negatives (those who fail to meet the threshold on the selection-context individual characteristic(s) but would have surpassed the acceptability threshold on job performance had they been hired). In practice, the number of false positives are seldom systematically tracked, and the number of false negatives cannot be

Stefan T. Mol - 9781803921761
Downloaded from https://www.elgaronline.com/ at 11/11/2024 03:41:49PM
via Open Access. This work is licensed under the Creative Commons
Attribution-NonCommercial-No Derivatives 4.0 License
https://creativecommons.org/licenses/by-nc-nd/4.0/

counted, because the performance of those who are not hired is not known. Sticking to the theory, however, the organization's interests are mainly aligned with preventing false positives, whereas the candidates' interests are aligned with a minimization of false negatives (particularly insofar as they turn out to be one!). This is noteworthy, because organizations may reduce the number of false positives at the cost of increasing the number of false negatives. That is (and this works better the larger the number of applicants), by only selecting the "crème de la crème" (i.e. setting the threshold to only include the very top scorers on the selection measure of relevant individual differences), the organization increases the odds of hiring the most capable candidate(s) while now incurring the cost of rejecting more candidates who would have been qualified.

Rather than raising the bar on whatever imperfect selection method we have chosen, an arguably more ethical approach to developing an evidence-based personnel selection system is to seek to optimize the prediction of future job performance, by optimizing criterion-related validity; that is, the empirically observed relationship between some measure of a (set of) predictor(s) and some measure of the criterion (i.e. job performance). Specifically, given the potentially vast personal consequences (e.g. unemployment) associated with candidates being rejected for a particular position, selection-context decision makers should strive on an ongoing basis to identify those individual characteristics that together provide the best possible prediction of future job performance. Interestingly, better prediction will simultaneously reduce the occurrence of both false positives and false negatives, to the point where (in theory at least) one could perfectly predict any candidates' performance, based solely on knowledge of the predictor scores. In measurement theory, when we observe such perfect correlation between two variables, we are inclined to conclude that we have measured the same thing twice. In personnel selection, we leverage the idea that past behavior is the best predictor of future behavior (Hough & Oswald, 2023) through the work sample test, a predictor measure designed to directly assess the performance domain by having applicants perform (parts of) the job for which they have been recruited.

One could be inclined to conclude then that personnel selection is solely about maximizing criterion-related validity, but that would be a gross oversimplification of matters. In a much recommended, albeit highly technical article, Binning and Barrett (1989) present some 19 distinct inferences that may jointly be used to support the validity of personnel decisions. Although it is beyond the scope of this entry to review their work in much detail, some fundamental points they make regarding the validity of selection measures are that: (i) actual job demands must be adequately represented in the performance domain through the process of job description, (ii) relying on criterion-related validity to evaluate the overall validity of the predictor is only justified insofar as both the predictor and criterion measure are construct-valid measures of the underlying (and theoretical) psychological construct and performance domains, and (iii) the choice for the specific predictor and criterion constructs must be justified on theoretical grounds. In this sense, criterion-related validity can be thought of as a necessary but insufficient condition to justify the validity of a predictor measure, and as an empirical proxy for the theoretical construct–construct relationships (i.e. those between the relatively stable individual characteristics and job performance) that lie at the heart of any personnel selection system. Indeed, absence of criterion-related validity for a particular predictor could be caused by any or all of the following: poor theory (i.e. selecting a predictor construct that has no bearing on the performance domain), poor operationalization of predictor and criterion measures (e.g. deficient or contaminated criterion measures) or sampling error (e.g. due to range restriction caused by validating the predictor(s) on a preselected sample of incumbents – who supposedly all surpass some minimal level of performance). Conversely, high criterion-related validity is no guarantee that the predictor will have utility in practice; for instance, when there is an incongruence between the delineated performance domain and the actual demands of the job. With regard to the latter, Binning and Barrett (1989, p. 489) notably state that the behavioral universes associated with meeting job demands have fuzzy boundaries, making their conclusive identification "logically impossible".

Job performance is typically (see, for instance, Motowidlo, 2003) defined in terms

Stefan T. Mol - 9781803921761
Downloaded from https://www.elgaronline.com/ at 11/17/2024 03:41:49PM
via Open Access. This work is licensed under the Creative Commons
Attribution-NonCommercial-No Derivatives 4.0 License
https://creativecommons.org/licenses/by-nc-nd/4.0/

of a theoretical construct domain that pertains to the value of the totality of individual job-related behaviors and outcomes that an individual employee contributes to the organization over a given time period. Relatedly, Thorndike (1949) coined the term "ultimate criterion" as denoting the complete domain of performance, which includes everything that ultimately defines success on the job. One question that the definition of job performance raises is what the appropriate time period ought to be. Thorndike's ultimate criterion implies that a comprehensive criterion ought to define and sample all behaviors and outcomes that a given job holder will contribute to their organization over the course of their tenure in that job, something that is clearly impractical. To summarize, the practical need to ultimately express the job performance of a sample of individuals on a single time-limited dimension is difficult to reconcile with the theoretical need to ensure that the sample of job performance that is ultimately assessed by the criterion measure is representative for the universe of behaviors and outcomes that comprise the "ultimate" job performance domain. Specification of the entire job performance domain through job description is the only means of ensuring such representativeness, in a manner analogous to how we can only evaluate the representativeness of a sample of individuals that is drawn from a population by knowing all individual members of that population. Indeed, Austin and Villanova (1992, p. 838) have defined the criterion as "a sample of performance (including behavior and outcomes), measured directly or indirectly, perceived to be of value to organizational constituencies for facilitating decisions about predictors or programs". In this context it is disappointing that we have yet to define a finite lingua franca with which all behaviors and outcomes in all jobs can be systematically described (although O*NET Online by the United States Department of Labor and the European Skills, Competences, Qualifications and Occupations classification [ESCO] are important developments to this end).

Another issue in the delineation of the performance domain that may be difficult to reconcile is that there may be different legitimate organizational stakeholders regarding an individual employee's job performance and each may value different, and at times even incompatible, behaviors and or outcomes (think, for example, of whistleblowing). It would seem that in these times of scarcity of talent and widespread occupational health issues, employees themselves are also an increasingly important stakeholder to the selection decision. Indeed, when we consider the duty of care that organizations have towards their employees, or when at least we abandon the notion of the employee as an expendable resource and consider the longer-term sustainability of employment and the increasing need for organizations to retain employees, it may start to make sense to expand the performance domain, and indeed the very definition of job performance, to include behaviors and outcomes that are valuable not solely to the organization but also to the individual. Increasingly, and in addition to just hiring for talent, organizations may need to distinguish themselves from their competitors by demonstrating that they facilitate well-being, growth, and personal development.

There is a pervasive tendency among personnel selection practitioners and researchers alike (myself included) to put the cart before the horse by mysteriously being lured into trying to answer the question of what it is that we should select on, before answering the question of what it is that we should select for. In this context, Jenkins (1946, p. 93) has famously remarked that, supposedly after predictor measures have been developed, criteria are "either given of god, or just found to be lying about". It cannot be emphasized enough that any inferentially valid approach to personnel selection must start by operationalizing the performance domain, or, stated differently, by developing a construct-valid criterion measure that can then be used as a target against which to evaluate the (battery of) construct-valid predictor measure(s). Once this target has been defined, criterion-related validity can then be established by examining the degree to which a particular predictor measure can explain variance in the criterion. In an effort to avoid the identification of spurious relationships between predictor measures and criterion measures, we must not only demonstrate that the predictor is (causally) related to the criterion, but also have some theoretical understanding of the process underlying this relationship. Such causal relationships are oftentimes difficult to evidence in practice because many of the individual differences that are studied in this context (other than job knowledge perhaps) tend to be

Stefan T. Mol - 9781803921761
Downloaded from https://www.elgaronline.com/ at 11/11/2024 03:41:49PM
via Open Access. This work is licensed under the Creative Commons
Attribution-NonCommercial-No Derivatives 4.0 License
https://creativecommons.org/licenses/by-nc-nd/4.0/

relatively enduring characteristics of people that are not amenable to the intervention that the generation of internal validity evidence would require.

Personnel selection, perhaps in part due to the relatively high impact of a number of now classic meta-analyses being published in this field (see, for instance, Barrick & Mount, 1991; Schmidt & Hunter, 1998), has also been characterized by a tendency to favor general as opposed to specific predictor constructs, although it appears such early meta-analytic work may have overcorrected for restriction of range, thereby overestimating predictive validity coefficients (Sackett et al., 2022). Despite this, the general consensus appears to be that structured interviews, job knowledge tests, biodata, work sample tests and cognitive ability tests are among the strongest predictors of job performance. What is all too easily forgotten in building on this work is that these meta-analyses average correlation coefficients across (for instance) different structured employment interviews for different jobs, in different organizations, in different sectors, at different job levels and across different countries, with oftentimes nontrivial variance in such coefficients being attributed to moderators that have yet to be identified. In this context it is telling that Sackett et al., (2022, p. 2063) conclude that the variability of validity estimates across settings "is an essential reminder that a given employer cannot count on the mean value as applicable to their organization". In my mind, this generalist approach has also resulted in a focus on broad constructs such as general mental ability and the general factor of personality that have thwarted theory development. What is more, many of the best predictors of job performance, and especially those that are more cognitively loaded, appear to be associated with an adverse impact on minorities, resulting in the so-called validity–diversity dilemma (see Rupp et al., 2020 for a potential solution).

In light of the aforementioned complexities associated with devising a valid personnel selection system (and many others not discussed here), it is perhaps not surprising that the systematic validation of the inferences underlying personnel selection in practice remains limited, and that there remains a significant divide between the science and practice of personnel selection. In redressing this state of affairs, it will likely be important to move away from a focus on (general) univariate predictors and criteria (Murphy, 2023), to a more comprehensive effort at fine-grained theory building, which ultimately will involve reviving the field of job analysis (Sanchez & Levine, 2012), and seeking to understand how jobs and occupations may be disaggregated into atomic yet meaningful parts (whether expressed in terms of universes of tasks, knowledge areas, skills or competencies), before examining why and how certain specific individual differences are antecedents of performance on those parts. Clearly such understanding would be invaluable, not only to justify personnel selection decisions, but also to identify idiosyncratic training needs in case no qualified applicant can be found.

In sum, the practical and intellectual challenges involved moving the field in this direction are daunting. I suspect, however, that by leveraging artificial intelligence, and more specifically generative-artificial-intelligenc e-based large language models (such as ChatGPT) we will enhance our ability to (pre) process the wealth of job-related information (obtained, for instance, from corpora of vacancies, resumes, technical manuals and/or online communications) and generate much more fine-grained "fledgling theories" than has been possible to date. What remains to be seen is whether such theories will ultimately be able to elucidate the causal mechanisms by means of which specific individual characteristics drive on-the-job behavior and outcomes. Clearly these theories too would still need to be validated (or refuted) along the lines that Binning and Barrett (1989) originally envisaged.

Stefan T. Mol

References

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology*, *77*(6), 836–874.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*(3), 478–494. https://doi.org/10.1037/0021-9010.74.3.478

Hough, L. M., & Oswald, F. L. (2023). Revisiting predictor–criterion construct congruence: Implications for designing personnel selec-

Stefan T. Mol - 9781803921761
Downloaded from https://www.elgaronline.com/ at 11/17/2024 03:41:49PM
via Open Access. This work is licensed under the Creative Commons
Attribution-NonCommercial-No Derivatives 4.0 License
https://creativecommons.org/licenses/by-nc-nd/4.0/

tion systems. *Industrial and Organizational Psychology*, *16*(3), 307–312.
Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology*, *10*(2), 93–98.
Motowidlo, S. J. (2003). Job performance. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology*, Vol. 12, pp. 39–53. John Wiley & Sons.
Murphy, K. (2023). Interpreting validity evidence: It is time to end the horse race. *Industrial and Organizational Psychology*, *16*(3), 341–343.
Roe, R. A. (2013). Personnel selection: Principles, models and techniques. In C. De Wolff, P. J. D. Drenth. & T. Henk (Eds.), *A Handbook of Work and Organizational Psychology, Volume 3: Personnel Psychology* (3rd ed., pp. 5–32). Psychology Press.
Rupp, D. E., Song, Q. C., & Strah, N. (2020). Addressing the so-called validity–diversity trade-off: Exploring the practicalities and legal defensibility of Pareto-optimization for reducing adverse impact within personnel selection. *Industrial and Organizational Psychology*, *13*(2), 246–271.
Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994
Sanchez, J. I., & Levine, E. L. (2012). The rise and fall of job analysis and the future of work analysis. *Annual Review of Psychology*, *63*(1), 397–425. https://doi.org/10.1146/annurev-psych-120710-100401
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*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262
Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. Wiley.

Stefan T. Mol - 9781803921761
Downloaded from https://www.elgaronline.com/ at 11/11/2024 03:41:49PM
via Open Access. This work is licensed under the Creative Commons
Attribution-NonCommercial-No Derivatives 4.0 License
https://creativecommons.org/licenses/by-nc-nd/4.0/