COMMENTARY

It takes more than meta-analysis to kill cognitive ability

Konrad Kulikowski

Institute of Management, Lodz University of Technology, Lodz, Poland
Email: konrad.kulikowski@p.lodz.pl

I would like to present methodological, theoretical, and practical arguments which suggest that Sackett et al. (2022, 2023) calls for revisiting the role of GMA in personnel selection are premature.

Methodology: Too many subjective decisions

Methodological concerns related to Sackett et al. (2022, 2023) meta-analysis were already raised by Oh, Le, and Roth (in press). I would like to concentrate here on the conceptual discussion of the two focal methodological decisions of Sackett et al., which, in my view, results in the lowering of the general mental ability (GMA) validity estimates. First, to not attempt a correction for range restriction and, second, to not control for job complexity.

Sackett et al. (2022, 2023) propose that range restriction is an issue mainly if the predictor in question was used in selecting the validation sample, and they argue that this would virtually never be the case in validation studies on GMA included in their meta-analysis, as this is unlikely that the same or similar GMA test scores used in validation procedures were also a part of the selection process. But, GMA was and still is considered one of the most important job performance predictors, as Kuncel, Ones, and Sackett (2010, p. 333) notice: "Cognitive ability is the workhorse of employee selection." So why it is unlikely that GMA tests of some sort were used as a basis of selection procedures in validation studies? This is unclear to me and not explained sufficiently. But even if we assume that GMA tests were never used as a part of selection procedures in studies analyzed by Sackett et al. (2022, 2023), this still does not mean that employees were not directly and indirectly selected based on GMA. It seems that Sackett et al. (2022) misinterpret the GMA test score with GMA as a construct itself, but it is not cognitive ability *tests* that predict performance but cognitive ability. It is important to notice that various GMA tests scores to a large extend reflect a common construct (Johnson et al., 2008), which "involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson, 1997a, p.13). This seems to be an ability developed in a time of the test, not the genetic potential or hereditary talent (see Schmidt, 2002). Assuming that it is unlikely that employees are directly hired based on their GMA is as assuming that it is unlikely that during selection procedures the abilities to plan, solve problems, reason, and learn from experience are directly taken into account by employers, but if these are not the basis for selection in most jobs, then what is? I argue that because of GMA's nature and importance in our lives (Brown et al., 2021), GMA as a construct is a vital criterion for selecting employees for jobs even if there is no formal GMA test in recruitment procedures. Jobs vary in cognitive complexity (https://www.onetonline.org/find/descriptor/browse/1.A), and job incumbents are not randomly assigned to jobs, but applicants take different jobs and occupy different positions based on the fit between their cognitive abilities and the complexity of the job (Gottfredson, 1997b). Individuals gravitate to jobs that are congruent with their cognitive ability (Judge et al., 2010; Wilk & Sackett, 1996) thus if

we test current employers in any occupation, this is a sample directly selected for the job based on their GMA, in such a way that GMA tends to fit tasks complexity level in a job (Gottfredson, 2002). Therefore GMA variation among given occupation incumbents is artificially lowered in comparison to a sample of applicants and even more in comparison to the general population because the selection on the job complexity directly reflected GMA.

Moreover, GMA is associated with numerous important occupational, educational, health, and social outcomes (Brown et al., 2021), so, there are many sources for indirect range restrictions. To mention three, GMA is reflected in education (Ritchie & Tucker-Drob, 2018), education might be a proxy for cognitive ability (Berry et al., 2006), and education often is a job-selection criterion. Second, job interviews of various forms are another source of indirect range restriction on GMA, as these complex social interactions correlate with GMA (Roth & Huffcutt, 2013), thus selecting based on an interview we indirectly select on GMA, and in most selection procedures, some form of interview is present. Third, GMA is related to emotional intelligence as the ability to understand and then regulate emotions (Joseph & Newman, 2010), which might be vital in making a positive impression on recruiters and further employers in selection processes. Besides that, there might be even more sources of range restriction as GMA is related to many important life outcomes such as income, leadership, unemployment, and physical and mental health that all might play a role of criterion in personnel selection. Thus, on a conceptual level, I see the Sackett et al. (2022) decision to avoid corrections for range restriction as unjustified. It is good to remain a wise warning "*Failure to take range restriction into account can dramatically distort research findings*" (Sackett et al., 2008, p. 217).

Second, it is of note that Schmidt and Hunter's (1998) findings about GMA critiqued by Sackett et al. (2022) related to jobs with medium complexity, whereas Sackett et al. (2022) do not control for job complexity as a moderator. This is surprising, as job complexity has an important effect on GMA validity for job performance when GMA validity increases as job complexity increases (Gottfredson, 1997b, 2002). Thus merging jobs of different complexity might give a blurred picture of GMA's practical validity.

One of the main suggestions of Sackett et al. (2022), is the rejection of Schmidt and Hunter's (1998) conclusion about GMA role in predicting performance claiming that "Cognitive ability is no longer the stand-out predictor that it was in the prior work. Structured interviews emerged as the predictor with the highest mean validity" (Sackett et al., 2023, p.5), and they asked: "So how did this happen?" My answer is this: They decide not to use correction for range restriction and do not control for job complexity. In my opinion, contrasting Schmidt and Hunter's (1998) meta-analytical estimate for GMA in medium complex jobs and with range restriction correction to Sackett et al. (2022) meta-analytical estimate without taking into account job complexity and with no range restrictions adjustments, is debatable as it is comparing apples and oranges.

Theory: Is meta-analysis a substitute for reasoning?

But even if we assume that the Sackett et al. (2022) meta-analysis is without any flaws, it is still important to note that Schmidt and Hunter (1998, 2004) positioned cognitive abilities as the most important predictor of job performance based not solely on meta-analysis estimates but also on theoretical and practical reasoning. Interestingly in Schmidt and Hunter's (1998) meta-analysis the highest validity in predicting job performance was found for the work sample tests but not for GMA, and structured employment interviews have as high validity as the GMA test (this can be also seen in Sackett et al., 2022, table 3). Schmidt and Hunter (1998) do not focus on GMA only due to high meta-analytical estimates but because all other personnel selection measures fall short in comparison to GMA when we consider a wider theoretical context. Thus the framing of Sackett et al. (2022) paper as an attempt to refute Schmidt and Hunter (1998) based on comparing meta-analytical estimates seems to me like a straw man argument. Meta-analytical estimates are not a

substitute for scientific reasoning and have no magical power to solve all controversies and pass a final verdict (Vrieze, 2018). Schmidt and Hunter (1998) suggest that GMA advantages come from many sources, not only from the fact that they have meta-analytical validity estimates higher by a few hundredths than other predictors. The GMA tests are valid, reliable, and have a clearly defined nomological network, whereas it is debatable even what is measured by some other personnel selection predictors. For example: What is measured by employment interviews: personality, GMA, integrity, knowledge, all of these or something else? Also, the evidence of GMA validity comes from many research but for other predictors, there is usually a relatively lower number of studies available, thus taking into account the crisis of reliability in psychological literature, their conclusions are less robust (e.g., total $k = 884$ and $N = 88\ 894$ for GMA vs $k = 105$ and $N = 7864$ for structured interviews, in Sackett et al., 2022, table 2). More importantly, Schmidt and Hunter (1998, 2004) highlight that GMA predicts not only job performance but also job-related learning and training performance (discussion on this is missing in Sackett et al., 2022). Further, GMA not only predicts but has a causal impact on performance, people with higher GMA acquire more job knowledge and acquire it faster, thus doing their job better (Schmidt, 2002; Schmidt, & Hunter, 2004). Whereas for many other predictors, mechanisms of its impact on performance are unknown or unclear; for example, what is the mechanism of which structural interviews influence job performance?

Practical and economical factors: A bird in the hand is worth two in the bush

Besides all the theoretical advantages that stem from more than 100 years of research on GMA that "amass a coherent body of empirical knowledge withstanding the test of time" (Lubinski, 2004, p.96), the second line of cognitive abilities defences lies in practical and economical factors. In applied business settings, it is not only the validity of the predictor but the efficiency that counts —the validity of the procedure in relation to the costs of its application. The personnel selection method with perfect validity estimates is of no use to me if I cannot afford to use it, and for the GMA tests, the validity to cost ratio seems to be still very good and better than, for example, employment interviews. GMA measures are ready to use standardized psychometric solutions of common construct, with proof of validity and reliability, in contrast to interviews, which are an umbrella term encompassing many different types and procedures of interviews (not only structured and unstructured; Huffcutt et al., 2014). Sackett et al., highlight the practical implications of their findings, but as there is no one standard "structural interview," then what practical advice can Maria, a manager of my neighborhood company, take from the conclusion, that structured interviews are valid predictors of job performance? Does this mean that simply arranging a set of random questions in a structuralized order is enough to predict performance? Possibly not. This means that Maria needs to invest time and money in developing valid, reliable, and fair employment interviews when on a shelf she has a ready-to-use GMA test. Then the GMA test might be applied to many jobs from entry to managerial level and from simple to complex, whereas structural interviews must be adapted to given positions and occupations and might be inappropriate in many selection contexts, for example when candidates lack relevant job knowledge. GMA tests might be also conducted even by inexperienced recruiters but interview often demands from interviewers not only substantial knowledge in the interview domain but also self-discipline and reflexivity about their prejudices to conduct them fairly, validly, and reliably. Despite all these efforts, it is not uncommon for two interviewers to come to different conclusions about the same applicant. In light of the many errors and cognitive biases (Kahneman, 2012, in human judgment (including interviewers), the GMA test seems to be not only cheaper but also fairer compared to interviews. Let's consider two applicants with the same abilities to reason, learn, and solve complex problems. During interviews structured or not, there is more room for subjective judgments and unfair evaluations of their abilities due to common cognitive biases

(e.g., halo effect, attribution bias, confirmation bias) or bigotry (e.g., stereotyping, prejudices) compared to when we use objective, standardized tests. Furthermore, candidates' perception of their ability is often only loosely related to their actual ability level (Freund & Kasten, 2012). Thus, avoiding the GMA test might favor not those with higher ability but those with higher self-confidence. Moreover, despite the "bad press" that cognitive tests have received, there is still a lack of robust evidence that the GMA test underpredicts job performance for racial or ethnic subgroups (see e.g., Sackett, et al., 2023), generally, applicants with similar GMA test scores have a similar level of job performance (Oh, 2022). Structurization of the interview might do a good job of diminishing various biases but it does not reduce them all. In my view, the higher the probability of facing discrimination, the higher the probability that objective criteria created by GMA will provide more equal opportunities. This is important in the light of presumed adverse impacts mentioned by Sackett et al., because in this context we should also consider the costs and side effects of rejecting standardized GMA tests in personnel selection (Oh, 2022).

Final thoughts

To sum it up, even in the absence of all methodological concerns to their meta-analysis, if Sackett et al. (2022) want to dethrone GMA and instead position structured interviews as the focal predictor in personnel selection they should provide a sound theoretical explanation and practical benefits for this move not only meta-analytical estimates, estimates that in my view, depend on too many researchers degrees of freedom. But as GMA is an important performance predictor, this does not mean that we should use only the GMA tests or that GMA is the only number that matters. It seems to me that on a practical level, Sackett et al. (2022) results might be seen as confirming old truth, we should use composite measures consisting of cognitive and noncognitive predictors to increase validity and utility, and reduce adverse impact in personnel selection (Schmidt, 2002). It is very difficult to predict performance in a job; thus, to meet the demands of real life, we need all hands on the deck, not academic fights whose predictor is the best.

References

Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, **91**(3), 696–705. https://doi.org/10.1037/0021-9010.91.3.696

Brown, M. I., Wai, J., & Chabris, C. F. (2021). Can you ever be too smart for your own good? Comparing linear and nonlinear effects of cognitive ability on life outcomes. *Perspectives on Psychological Science*, **16**(6), 1337–1359. https://doi.org/10.1177/1745691620964122

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, **138**(2), 296–321. https://doi.org/10.1037/a0026556

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, **24**(1), 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence*, **24**(1), 79–132.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, **15**(1-2), 25–46. https://doi.org/10.1016/S0160-2896(97)90014-3

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment*, **22**(3), 297–309. https://doi.org/10.1111/ijsa.12078

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. Jr (2008). Still just 1 *g*: Consistent results from five test batteries. *Intelligence*, **36**(1), 81–95. https://doi.org/10.1016/j.intell.2007.06.001

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, **95**(1), 54–78. https://doi.org/10.1037/a0017286

Judge, T. A., Klinger, R. L., & Simon, L. S. (2010). Time is on my side: Time, general mental ability, human capital, and extrinsic career success. *Journal of Applied Psychology*, **95**(1), 92–107. https://doi.org/10.1037/a0017594

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Random House UK.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences*, **49**(4), 331–336. https://doi.org/10.1016/j.paid.2010.03.042

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, **86**(1), 96–111. https://doi.org/10.1037/0022-3514.86.1.96

Oh, I., Le, H., & Roth, P. L (in press). Revisiting Sackett et al.'s (2022) recommendation against correcting for range restriction in concurrent validation studies. *Journal of Applied Psychology*. https://doi.org/10.2139/ssrn.4308528

Oh, I. S. (2022). Perfect is the enemy of good enough: Putting the side effects of intelligence testing in perspective. *Industrial and Organizational Psychology*, **15**(1), 130–134. https://doi.org/10.1017/iop.2021.126

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, **29**(8), 1358–1369. https://doi.org/10.1177/0956797618774253

Roth, P. L., & Huffcutt, A. I. (2013). A meta-analysis of interviews and cognitive ability. *Journal of Personnel Psychology*, **2**(4), 157–169. https://doi.org/10.1027/1866-5888/a000091

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, **63**(4), 215–227. https://doi.org/10.1037/0003-066X.63.4.215

Sackett, P. R., Zhang, C., & Berry, C. M. (2023). Challenging conclusions about predictive bias against Hispanic test takers in personnel selection. *Journal of Applied Psychology*, **108**(2), 341–349. https://doi.org/10.1037/apl0000978

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. Retrieved from https://ink.library.smu.edu.sg/lkcsb_research/6894

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **16**(3), 283–300. https://doi.org/10.1017/iop.2023.24

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, **15**(1-2), 187–211. https://doi.org/10.1207/S15327043HUP1501&02_12

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, **86**(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

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

Vrieze, J. (2018, September 18). Meta-analyses were supposed to end scientific debates. Often, they only cause more controversy. *Science*. https://doi.org/10.1126/science.aav4617

Wilk, S. L., & Sackett, P. R. (1996). Longitudinal analysis of ability–job complexity fit and job change. *Personnel Psychology*, **49**(4), 937–967. https://doi.org/10.1111/j.1744-6570.1996.tb02455.x