

## Practice and retest effects in longitudinal studies of cognitive functioning

In this issue, Goldberg et al. [1] draw attention to a critically important aspect of studies of cognitive change: practice and retest effects in repeatedly observed cognitive test performance. Practice and retest effects are large, pervasive, and underappreciated. By large, I mean that average gains on repeat administration are often much greater than normative cognitive change over a similar interval [2]. By pervasive, I mean that practice and retest effects are seen for a wide variety of cognitive tests assessing different domains of functioning [3]. By underappreciated, I mean that despite a long-standing and enduring literature [4,5], investigators continue to develop protocols that reveal a failure to consider the full range of impact of practice and retest effects [6].

Goldberg et al. focus their review on the potential impact of practice and retest effects in randomized controlled trials in preclinical Alzheimer's disease. Of the range of potential uses of serial cognitive assessment—including clinical practice and observational and natural history studies—one might assume that randomized controlled trials may be a context in which practice and retest effects are of the least concern. This is because, as the thinking goes, even if practice and retest effects are observed, they should be present in equal measure in our randomly assigned control and active treatment groups and cancel out in group comparisons of treatment effects. Goldberg et al. argue that the assumption that practice and retest effects are equivalent in treated and controls may not be justified, and moreover that practice and retest effects may result from cognitive processes that are potentially distinct from the one intended to be measured with a given test. This is potentially important for emerging studies of disease-modifying therapies for Alzheimer's disease.

Perhaps, one of the reasons why some studies fail to plan for practice and retest effects in design and/or analysis is the lack of consensus on best methods. Thorndike [4] suggested practice and retest effects could be eliminated with 10 minutes of practice (similar to the massed practice approach by Goldberg et al.), but this does not eliminate practice and retest effects from performance, only from collected data. If practice and retest effects have inferential value (c.f., [7]), then this approach is not useful. Goldberg et al. recommend the use of alternate forms but also note critical

limitations of this approach. The literature suggests that alternate forms may attenuate but do not eliminate practice and retest effects [8]. If forms are not psychometrically equivalent and are not administered in a counterbalanced and randomized order, the use of alternate forms can introduce as much construct-irrelevant variance as practice and retest effects [6,9]. As Goldberg et al. mention, reliable change indices (RCIs), with correction for practice [10], have a certain appeal. However, early proponents of RCIs now favor regression-based approaches [11], and not all authors have found practice-corrected RCIs to be useful in the context of randomized controlled trials [12].

Another approach to deal with practice and retest effects not mentioned by Goldberg et al. is statistical modeling in repeated-measures designs with linear mixed effect or related data analysis approaches [13,14]. Hoffman et al. [15] offer the important caveat that in typical fixed-interval designs involving age-heterogeneous samples, such approaches are in general not informative about retest effects because of the confounding of age differences and retest-related gains [15]. These authors suggest instead a seldom-used approach to assessment and modeling of cognitive performance data: the “measurement burst” design [16]. The goal of such designs is to model individual variability in repeat administration over a short enough interval to render aging effects negligible and model aging trends with repeated short-interval bursts of measurement over longer intervals. Salthouse [17] has used such approaches in the study of cognitive performance and revealed important and deeper complexities associated with the practice and retest effect. Not only can we expect age differences in the magnitude of retest-related gains (which may be of clinical relevance [7]), but the retention of retest-related gains over longer intervals is also revealed as a potentially important individual difference factor and one that confounds but is not conceptually equivalent to aging-related change in cognitive ability.

Despite a century of research, there is no clear consensus on the best methods to address the (still-emerging) impact of practice and retest effects. Nevertheless, recommendation by Goldberg et al. to use tests designed to minimize practice and retest effects without changing the fundamental construct of interest is a good one. Investigators should be wary of

parallel-but-not-equivalent alternate forms and designs that do not include counterbalancing [6]. The development of cognitive tests calibrated with item response theory methods and administered in an adaptive fashion, such as the NIH Toolbox [18], offers great potential in this regard. In computerized adaptive testing, psychometric equivalence and effective counterbalancing can be incorporated into the adaptive testing procedure. A plan for using design and analysis features (e.g., control or comparison groups, measurement burst design) to adjust and/or model the impact of practice and retest effects is now something that reviewers should expect to see in research proposals.

Richard N. Jones\*

Senior Associate Editor  
Alzheimer's & Dementia: Diagnosis  
Assessment & Disease Monitoring

Director  
Quantitative Science Program

Associate Professor  
Department of Psychiatry and Human Behavior  
Warren Alpert Medical School  
Brown University  
345 Blackstone Boulevard, Box G-BH  
Providence, RI 02906-4800, USA

\*Corresponding author. Tel.: 401-444-1943;  
Fax: 401-444-1911.

*E-mail address:* [Richard\\_Jones@Brown.edu](mailto:Richard_Jones@Brown.edu)

## References

- [1] Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *DADM* 2015. in press.
- [2] Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci* 2010;11:118.
- [3] McCaffrey RJ, Westervelt HJ. Issues associated with repeated neuropsychological assessments. *Neuropsychol Rev* 1995;5:203–21.
- [4] Thorndike EL. Practice effects in intelligence tests. *J Exp Psychol* 1922;5:101.
- [5] Salthouse T. Frequent assessments may obscure cognitive decline. *Psychol Assess* 2014;26:1063.
- [6] Gross AL, Inouye SK, Rebok GW, Brandt J, Crane PK, Parisi JM, et al. Parallel but not equivalent: challenges and solutions for repeated assessment of cognition over time. *J Clin Exp Neuropsychol* 2012;34:758–72.
- [7] Mormino EC, Betensky RA, Hedden T, Schultz AP, Amariglio RE, Rentz DM, et al. Synergistic effect of  $\beta$ -amyloid and neurodegeneration on cognitive decline in clinically normal individuals. *JAMA Neurol* 2014;71:1379–85.
- [8] Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. *Arch Clin Neuropsychol* 2005;20:517–29.
- [9] Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol* 1998;20:339–52.
- [10] Iverson GL. Interpreting change on the WAIS-III/WMS-III in clinical samples. *Arch Clin Neuropsychol* 2001;16:183–91.
- [11] Crawford JR, Garthwaite PH, Denham AK, Chelune GJ. Using regression equations built from summary data in the psychological assessment of the individual case: Extension to multiple regression. *Psychol Assess* 2012;24:801.
- [12] Di Bari M, Pahor M, Barnard M, Gades N, Graney M, Franse LV, et al. Evaluation and correction for a 'training effect' in the cognitive assessment of older adults. *Neuroepidemiology* 2002;21:87–92.
- [13] Wilson RS, Li Y, Bienias JL, Bennett DA. Cognitive decline in old age: Separating retest effects from the effects of growing older. *Psychol Aging* 2006;21:774.
- [14] Rabbitt P, Diggle P, Holland F, McInnes L. Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *J Gerontol B Psychol Sci Soc Sci* 2004;59:P84–97.
- [15] Hoffman L, Hofer SM, Sliwinski MJ. On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. *Psychol Aging* 2011;26:778.
- [16] Sliwinski MJ. Measurement-burst designs for social health research. *Soc Personal Psychol Compass* 2008;2:245–61.
- [17] Salthouse TA. Effects of age and ability on components of cognitive change. *Intelligence* 2013;41:501–11.
- [18] Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. *Neurology* 2013;80(11 Suppl 3):S54–64.