Trends in
Cognitive Sciences

CellPress

Science & Society

Can neuropsychological testing be improved with model-based approaches?

Roger Ratcliff [1,*] and Gail McKoon [1]

There has been little impact of cognitive psychology and modeling on neuropsychological testing for over 50 years. There is also a disconnect between those tests and the constructs they are said to measure. We discuss studies at the interface between testing and modeling that illustrate the opportunity for advances.

Neuropsychological testing is a massive industry that costs society hundreds of millions of dollars per year in the United States alone [1] with insurance paying for much of it. Tests are used to evaluate whether patients have deficits caused by Alzheimer's disease, attention deficit hyperactivity disorder (ADHD), concussion, schizophrenia, Parkinson's disease, chemotherapy, and so on. However, the tests currently in widespread clinical use (even in recent approaches [2]) have not been informed by results from cognitive research. This presents a compelling opportunity – to use modern cognitive and modeling research to investigate and perhaps develop neuropsychological tests.

In diagnostic tests, simple tasks are performed and the results are scored according to published norms. Most tests have high reliability and are argued to have high validity in measuring the constructs they are intended to evaluate. Many tests are commercial and obtaining the materials, scoring sheets, and norms requires a subscription or purchase. Individual testing is usually required, which can take several hours. Other commercial tests are computer based, with results transmitted back to the company for scoring and research purposes. The commercial aspect of many tests has likely contributed to resistance to change.

Neuropsychological tests are usually assumed to measure a specific ability or construct. The Compendium of Neuropsychological Tests [3] organizes tests in categories of executive functioning, attention, memory, and so on. These terms, as well as working memory, inhibitory control, and speed of processing, are used to suggest a theoretical understanding of common sense abilities, but to move beyond this informal organization, the terms require a detailed understanding of the representations and processes involved in them. Moreover, none of the constructs can really be considered independent of the others; executive function depends on attention to and information from working memory and on the speed with which processes operate before information is lost.

In practice, neuropsychological tests are usually used to support a diagnosis based on other instruments (interviews with a patient and caregiver, sometimes diagnostic tools such as neuroimaging). By themselves, neuropsychological tests can identify large deficits, but usually cannot identify subtle deficits (mild memory impairment, cognitive impairment under chemotherapy, and ADHD), the onset of a deficit, the rate of decline, or even whether decline will occur. Dissatisfaction with this situation has been apparent for some time [4,5], but one factor contributing to the lack of progress is likely the long history and large amount of norming data for tests, which makes abandoning or modifying them difficult.

An example of a target for new approaches is speed of processing, a construct thought to be a fundamental property of the information-processing system. It is part of the Wechsler Adult Intelligence Scales III and IV intelligence quotient tests and has been a component of a large cognitive training study (IMPACT). Several tasks are used to measure speed of processing, for example, trail-making and color–word interference (Stroop), but these are also said to measure attention [3] and executive function (the Delis–Kaplan Executive Function System), respectively. In general, there is little agreement about which tasks measure which constructs and there is no theory to specify which dependent variable within a task measures the construct. Furthermore, well-known tight relationships between accuracy and response times (RTs) [6] have been ignored in the speed of processing construct, and processes involved in the component tasks are not specified.

RT modeling has been highly active over the past 50 years in cognitive modeling, but there is a major disconnect with 'speed of processing'. Neither the review of the past 50 years of research [7] nor the edited book [8] on 'processing speed in clinical populations' makes any reference to recent modeling (Box 1 discusses RT modeling with application to aging research). This lack of impact is surprising, but there are structural reasons for it. Cognitive psychologists, particularly theorists, have not been motivated to translate their work into domains other than their own because of the high cost in time and resources required, as well as a historical lack of research funding from agencies that might want to see translation into clinical practices. Neuropsychologists have time constraints in their training that prevent them from acquiring the high degree of technical expertise and training needed to understand and use this approach.

The major problem in developing new tests is discovering tasks, manipulations,

Box 1. Model-based approaches to accuracy and RTs

Theory is needed to explain the relationships between accuracy and RTs and diffusion models [12] do this. By adjusting a model's parameters until its predictions match the data, the model takes accuracy and RTs for each subject and decomposes them into the processing components that underlie decisions. In diffusion models, noisy information from perception or memory is accumulated over time from a starting point until it reaches a criterion, at which point a decision is made. The model separates information that drives the accumulation process from settings of criteria and processing biases.

In recognition memory, long RTs suggest deficits with age, but accuracy suggests no deficit. Model analyses have shown that older adults adopt more conservative decision criteria than young adults, but the information they use in the decision process is similar, thus resolving the discrepancy [13].

To study individual differences, parameter estimates must be reliable. Although one article reported low test–retest reliability in diffusion model parameters [14], inadequate numbers of observations per subject per task (84 on average) were used. With enough observations (several hundred), not only are correlations of parameters within a task reliable, but correlations between tasks are as large as 0.7 [13].

and measures that target specific disorders. Box 2 shows an example of a diffusion decision model-based approach that accounts for accuracy and RTs in a memory task with memory-disordered patients and controls that produced discrimination between the two groups that was comparable to commercial products [9].

In contrast to the study in Box 1, model-based analyses can also have problems if performed poorly. Fifty studies of ADHD with 15 tasks using model-based analyses of RT and sometimes accuracy were reviewed in [10]. Each study specified constructs that it was intended to investigate: cognitive flexibility, sustained and selective attention, and so on, but the tasks, measures, constructs, and modeling of the data were so disjointed that there was no basis for drawing broad conclusions. Among the studies reviewed, tasks and participant groups differed so much that they could not be compared. Different results between studies could be attributed to differences in subject inclusion criteria or differences in the details of task procedures. In standard neuropsychological testing, great care is taken to standardize tests, but in the model-based studies reviewed, there was no such standardization. At the level of behavioral analyses, choices of dependent variables were inconsistent, with some studies using RTs, others accuracy, and others using both (the same inconsistency arises in numeracy research [11]). Tasks were also inconsistently assigned to constructs across the ADHD studies; Stroop and Navon tasks each were in three different constructs and flanker and continuous performance tasks were in two. Similarly, in seven meta-analyses of studies of cognitive impairments due to chemotherapy [4], every domain was impaired in at least one meta-analysis and unimpaired in at least one other, and single tasks were assigned to multiple domains.

Box 2. Model-based approaches to memory disorders

Memory-disordered patients and unimpaired subjects were tested on a simple memory task and the RT and accuracy data were fit with the diffusion model in [9]. The parameters obtained for two-thirds of the subjects and their clinical diagnoses were used to train a classifier (e.g., linear discriminant analysis) to obtain weights on the parameters that best discriminated between patients and controls. These weights were used to classify the other one-third of the subjects (cross validation that controls for overfitting). These classifications matched those of the diagnoses for 83% of the subjects, about the same as the best commercial products. Around 83% accuracy may be as good as possible because postmortem studies of patients diagnosed with Alzheimer's disease found clinical diagnoses to be only about 85–90% correct.

This model-based approach has an important validity check that standard neuropsychological tests do not: the model has to fit data adequately; if it does not, the model cannot be used. Besides collecting enough data to produce reliable parameter estimates [13], the differences between patients and controls must be large enough to be clinically meaningful. This means that the components of processing must target the disorder that is being evaluated, as for the task in Box 1.

For ADHD children, longer RTs might stem from deficits in the information driving a decision process. But they might also stem from an effort to slow responding to avoid errors. A model-based analysis can separate these two possibilities (Box 1).

All of these issues raise the question: what do neuropsychological tasks actually measure? Terms such as selective attention and inhibitory control might be useful for talking about deficits, but the ambiguity in the mapping from task to construct and the lack of understanding of how processes relate to performance measures make it impossible to unambiguously understand the locus of deficits. In addition, mapping from a single dependent variable onto a single ability is unreasonable.

Neural measures might serve as an alternative to neuropsychological testing, but they also have problems [10]. One is that constructs (such as inhibitory control) cannot specify which event-related potential components, which frequency bands, or at what time points these measures should be used to evaluate hypotheses about processing. Likewise, activity-based measures and neural theories based on resting state analyses do not provide testable, quantitative predictions because they do not map onto behavioral or clinical measures in any agreed-upon way. However, we argue that these inadequate links between theory and measures should be seen as a challenge to advance theory to impact clinical issues rather than a limitation.

There are promising signs of increased interest at NIH in computational and modern experimental approaches to clinical

disorders, and several programs support computational modeling with behavioral components such as theoretical and computational neuroscience and computational psychiatry. The National Cancer Institute, the National Institute of Child Health and Human Development, and the National Institute on Aging (NIA) have robust programs supporting behavioral research and the NIA recently published a call for proposals: 'Screening for Cognitive Impairment: Decision-Making', RFA-AG-23-007.

The challenge is to use the past 40 years of cognitive research to improve neuropsychological testing. Clearly, small tweaks to the existing approach will not do. To make progress, the hunt should be on for tasks and models of them that target individual differences in specific deficits. Care has to be taken in that clinical deficits may not match cognitive concepts that have the same name; for example, a clinical lack of inhibition may not be the same as brief inhibitory effects in a flanker task. Multiple tasks that target different aspects of a deficit should be sought because combinations of such tasks could produce better diagnostics. However, this research needs a significant commitment because controlled protocols are needed that use dozens or even a hundred or more clinically diagnosed patients and similar-sized control groups. In many cases, simply adding 10 min of a task to an ongoing protocol will not be sufficient. Small teams of clinicians and cognitive psychologists who understand each other's research and can work together are needed along with the resources that provide the numbers of patients and controls to explore tasks, measures, and models. This is an important opportunity that the field of cognitive psychology cannot afford to ignore.

Acknowledgments

This work was supported by funding from the National Institute on Aging (Grant numbers R01-AG041176 to R.R., principal investigator; and R01-AG057841 to R.R., principal investigator). We thank Todd Horowitz, Jerry Suls, and Luke Stoeckel for their comments on this article.

Declaration of interests

The authors have no interests to declare.

Resources

[i]https://healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox

[1]The Ohio State University, Columbus, OH, USA

*Correspondence:
ratcliff.22@osu.edu (R. Ratcliff).

https://doi.org/10.1016/j.tics.2022.08.015

References

1. Institute of Medicine (2015) *Psychological Testing in the Service of Disability Determination*, National Academies Press
2. Wefel, J.S. *et al.* (2011) International Cognition and Cancer Task Force recommendations to harmonise studies of cognitive function in patients with cancer. *Lancet Oncol.* 12, 703–708
3. Strauss, E. *et al.* (2006) *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary*, Oxford University Press, New York
4. Horowitz, T.S. *et al.* (2019) Understanding the profile of cancer-related cognitive impairments: a critique of meta-analyses. *J. Natl. Cancer Inst.* 111, 1009–1015
5. Nelson, W.L. and Suls, J. (2013) New approaches to understand cognitive changes associated with chemotherapy for non-central nervous system tumors. *J. Pain Symptom Manag.* 46, 707–721
6. Ratcliff, R. *et al.* (2015) Modeling regularities in response time and accuracy data with the diffusion model. *Curr. Dir. Psychol. Sci.* 24, 458–470
7. Sheppard, L.D. (2008) Intelligence and speed of information-processing: a review of 50 years of research. *Personal. Individ. Differ.* 44, 533–549
8. DeLuca, J., Kalmar, J.H., eds (2008) *Information Processing Speed in Clinical Populations*, Taylor & Francis
9. Ratcliff, R. *et al.* (2022) Discriminating memory disordered patients from controls using diffusion model parameters from recognition memory. *J. Exp. Psychol. Gen.* 151, 1377–1393
10. Ging-Jehli, N. *et al.* (2021) Improving neurocognitive testing using computational psychiatry - a systematic review for ADHD. *Psychol. Bull.* 147, 169–231
11. Ratcliff, R. *et al.* (2015) Modeling individual differences in response time and accuracy in numeracy. *Cognition* 137, 115–136
12. Ratcliff, R. and McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922
13. Ratcliff, R. *et al.* (2010) Individual differences, aging, and IQ in two-choice tasks. *Cogn. Psychol.* 60, 127–157
14. Enkavi, A.Z. *et al.* (2019) Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci. U. S. A.* 116, 5472–5477