THE CLINICAL NEUROPSYCHOLOGIST
2019, VOL. 33, NO. 2, 357–368
https://doi.org/10.1080/13854046.2018.1518489

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience

Roy P. C. Kessels[a,b,c]

[a]Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands; [b]Department of Medical Psychology, Radboud University Medical Center, Nijmegen, the Netherlands; [c]Vincent van Gogh Institute for Psychiatry, Centre of Excellence for Korsakoff and Alcohol-Related Cognitive Disorders, Venray, the Netherlands

ABSTRACT

Objective: In cognitive neuroscience, well-controlled and highly specific paradigms have been developed to measure cognitive processes over the last decades, often using computer-assisted presentation and response registration. This approach is in contrast with the traditional paper-and-pencil tests used in clinical neuropsychology, which typically assess cognitive function in a less specific manner, often even at the level of a cognitive domain. As a result, important aspects of cognitive (dys)function may be missed during a neuropsychological assessment. This paper focuses on the main challenges that need to be overcome in order to successfully integrate experimental paradigms from cognitive neuroscience into the clinical practice of neuropsychologists.

Method: Six challenges are discussed: (i) experimental paradigms are often lengthy and may be overly specific; (ii) technical limitations even today hamper their application in clinics; (iii) the psychometric properties of methods used in cognitive neuroscience are under-examined or poor; (iv) many paradigms from cognitive neuroscience rely on reaction times rather than accuracy, limiting their use in the many brain-injured patients with processing speed deficits; (v) the predictive and ecological validity of these paradigms often unclear; (vi) technological progress (e.g. Moore's law) seriously affects the continuous availability of experimental computerized assessment methods.

Conclusion: Both cognitive neuroscientists and clinical neuropsychologists should work together to develop and validate novel paradigms for use in clinical assessments that are platform-independent, reliable and valid, user friendly and easy to use in clinical practice.

ARTICLE HISTORY

Received 29 March 2018
Accepted 28 August 2018

KEYWORDS

Neuropsychological assessment; computer-assisted testing; cognitive neuroscience; psychometrics; precision neuropsychology; experience sampling; virtual reality

CONTACT Roy P. C. Kessels r.kessels@donders.ru.nl Radboud University, DCC – Neuropsychology and Rehabilitation Psychology, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Introduction

Neuropsychological assessment is aimed at characterizing individual differences in cognitive function, with the eventual goal to estimate the probability that a given performance is 'impaired', adjusted for confounding factors, such as age, gender, socio-economic status and ethnicity. Such an assessment typically results in a cognitive profile, in which strengths, weaknesses, and impairments in specific cognitive processes are identified which – for the sake of convenience – are usually grouped into cognitive domains. These cognitive profiles can subsequently be used for medical decision-making (e.g. from a diagnostic perspective), treatment identification or monitoring, and psychoeducation, taking into account the individual's premorbid cognitive status as well as specific information about the somatic status, brain disease or cerebral dysfunction. Neuropsychological assessment as a discipline that is part of clinical neuropsychology has become an established field, with its own research – for instance, on the psychometric characteristics of individual tests –, dedicated handbooks (e.g. *Neuropsychological Assessment* by Lezak, Howieson, Bigler & Tranel, 2012) and journals, specific clinical training programs aimed at assessment skills in settings varying from child neuropsychology, forensic settings, memory clinics, rehabilitation centers to general hospitals, and normative data sets for use individuals from all over the world (see, for instance, Strauss, Sherman, & Spreen, 2006).

In order to make reliable and valid claim about someone's cognitive profile, numerous paper-and-pencil neuropsychological tests have been developed, each with its own strengths and limitations, but when administered together, as part of an 'assessment'[1] are considered the gold standard for making such claims. These paper-and-pencil tests are broadly available, sometimes even in the public domain. Extensive normative data sets are available for most of these paper-and-pencil tests – although their quantity and quality may vary – and their psychometric properties have often been extensively studied (i.e. their validity and reliability; see e.g. Lezak et al., 2012). These tests have not only been validated in large samples of healthy individuals from ages that often cover the full lifespan, but also in clinical populations (notably patients with specific neurological diseases, such as Alzheimer's disease, stroke or traumatic brain injury, psychiatric disorders, ranging from Attention Deficit Hyperactivity Disorder (ADHD) and autism to schizophrenia and obsessive-compulsive disorder, and other somatic disorders that may or may not affect the brain directly, including human immunodeficiency virus, cancer and diabetes). Many of these tests have become the 'classics' of neuropsychological assessment, i.e. tests that are available in many languages and are widely used.

Interestingly, these classic paper-and-pencil tests have been around for decades, with some even over a hundred years old. For instance, the Digit Span paradigm was first described by Jacobs (1887), the Trail Making Test (TMT) originates from the Army Individual Test Battery published in the 1940s for use in the United States, Rey's complex figure was described in 1941 in André Rey's paper on the assessment of traumatic brain injury and the Stroop Color Word Test originates from experimental studies by Stroop (1935). Many of these classic tests have also been made available as computerized versions, typically mimicking the paper-and-pencil administration as much as

possible, albeit that the keyboard, mouse or touch-sensitive screen is used for measuring the responses (see e.g. Bauer, Iverson, Cernich, Binder, Ruff and Naugle, 2012).

Despite the fact that these classic paper-and-pencil tests are being widely used, this approach has several important limitations, which hamper the valid interpretation of the tests' results. First, given that many of these tests are old, the principles on which they are based are often outdated or may even have become obsolete. As a result, many of these tests are not considered to be process-pure from modern day's perspective. An example is the TMT, of which the first incarnation– The Taylor Number Series – even predates the TMT version included in the Army Individual Test of General Ability (Boake, 2002). It was developed as a performance test with the aim to assess an individual's intelligence, motor skills and alertness and later considered a test to measure 'organicity' (see Eling, 2013, for a detailed description of its origins). However, the TMT is a multifactorial test that is sensitive to deficits in working memory – and executive function in a broader sense, episodic memory, processing speed and attention (see e.g. Oosterman, Vogels, Van Harten et al., 2010). This makes specific claims about the underlying neurocognitive mechanisms of a poor performance on this test complicated, although the test can still be a valid instrument for distinguishing healthy individuals from brain-diseased patients. This lack of process-purity – also seen in other tests or batteries, such as the Behavioral Assessment of the Dysexecutive Syndrome (BADS), Rey's complex figure or the Wisconsin Card Sorting Test (see Lezak et al., 2012) – calls for novel paradigms that will facilitate 'precision neuropsychology', i.e. unraveling specific impairments in detailed cognitive processes rather than broad cognitive domains (or even 'cognition' in general).

Paradigms in cognitive neuroscience

Cognitive neuroscience (CNS), a term coined by Michael S. Gazzaniga, is the research field aimed at understanding 'how the functions of the physical brain can yield the thoughts and ideas of an intangible mind' (Gazzaniga, Ivry & Mangun, 2009, p. 3). This field has developed – and continues to develop – numerous paradigms that enable the precise, fine-grained unraveling of cognitive processing, often from a translational perspective, resulting in tasks that have an analogue for use in rodents, primates and humans. Many of these paradigms have been used in studies focusing on the neural underpinnings of the involved cognitive processes (for examples, using functional magnetic resonance imaging or electroencephalography) and many of these tasks have also been used in patient studies. Examples include Random Number Generation paradigms to assess specific cognitive control processes (e.g. Maes, Eling, Reelick & Kessels, 2011), *N*-back paradigms to study working memory updating in which the cognitive load can be systematically varied (Kirchner, 1958), the Iowa Gambling Task (IGT) to measure decision-making and risk taking (Bechara, Damasio, Tranel, & Damasio, 2005), Serial Reaction Time tasks for examining motor-sequence learning (Nissen & Bullemer, 1987) or reversal learning paradigms for assessing adaptive behavior (see Figure 1 for an example). The different outcome variables have often been linked to highly specific neurocognitive processes, often embedded in specific theories and models. Consequently, one could argue that it is these paradigms that could be

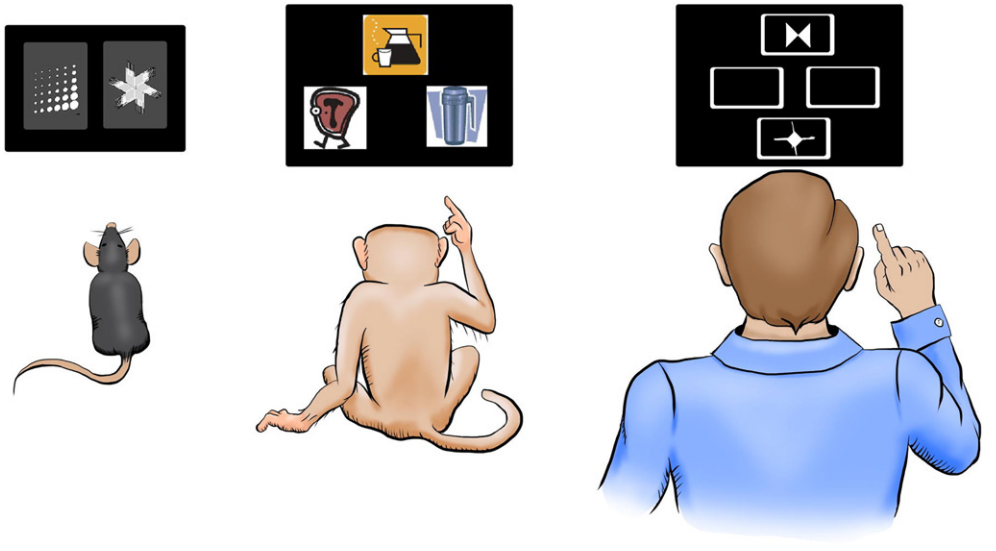


Figure 1. Example of a reversal learning paradigm that has been developed from a translational perspective. Rodents, primates or humans have to detect partially or fully predictive outcomes of stimuli and – once acquired – have to flexibly adapt to new stimulus-outcome contingencies. Reprinted from *Neuroscience*, 345, A. Izquierdo, J.L. Brigman, A.K. Radke, P.H. Rudebeck, & A. Holmes, The neural basis of reversal learning: An updated perspective, pp. 12–16, © 2017, with permission from Elsevier.

the ideal candidates for substituting the classic, process-impure neuropsychological tests discussed above. However, these often innovative paradigms are rarely used in clinical practice. Why is it that they do not find their way to the clinic? Here, I will present and discuss six challenges that must be overcome, which may guide test development for use in clinical assessments.

Challenge 1: experimental paradigms are often lengthy and may be overly specific

Consider the following case description:

> Mr. Jones is 68 years of age, highly educated, and visits a neurologist because he experienced multiple falls and has difficulty adjusting to new situations. He also experiences mental fatigue. His wife also mentions everyday cognitive problems, such as forgetfulness, difficulty acquiring new skills and planning difficulties. The neurologist refers Mr. Jones to a clinical neuropsychologist for a full assessment. The clinical neuropsychologist administers a combination of classic paper-and-pencil tasks and several computer tasks that are based on evidence from cognitive neuroscience. (see Table 1)

The examples in this proposed test battery illustrate this challenge. That is, while the paper-and-pencil tasks are short, with relatively few trials and only a few outcome measures, the computerized paradigms are lengthy, involve many trials and result in numerous outcome measures. Potentially, highly specific outcome measures enable a clinical interpretation that is also highly precise ('the patient has difficulty adjusting to new situations, which may be related to the inability to maintain target information in

Table 1. Example of a test battery for Mr. Jones, consisting of both classic paper-and-pencil tests (P&P) and paradigms originating from cognitive neuroscience (CNS).

Test	Cognitive process	P&P/CNS	Administration duration	No. of trials / items	No. of outcome variables
Wechsler spatial span	Working-memory updating	P&P	4 min	20–25	2
Trail making test	Cognitive flexibility/processing speed	P&P	6 min	3	3
Stroop color word test	Response inhibition	P&P	6 min	3	3
Probabilistic reversal learning	Adaptability	CNS	45 min	2 × 500	6
CANTAB® spatial working memory	Working-memory updating	CNS	8 min	50+	24
Serial reaction time task	Implicit motor learning	CNS	20 min	6 × 100	6
Montreal cognitive assessment	Overall 'cognition'	P&P	10 min	14	1
Brixton spatial anticipation test	Rule detection	P&P	7 min	56	1

working memory across trials, while having intact visuospatial sketchpad processing'). Such interpretation relies heavily on the expertise of the given clinical neuropsychologist though. Moreover, even if the task at hand has been validated and is reliable, its outcome measures – notably specific indices such as difference scores contrasting two conditions – may not always be. Furthermore, such a precise assessment may not always be helpful in clinical practice, as conclusions are often presented at the level of cognitive subdomains rather than based on highly specific cognitive processes that in many disorders are also not selectively affected. Additionally, multiple outcome measures may even complicate the clinical interpretation, for instance, in the case of conflicting results, and the cost–benefit ratio of administering lengthy, overly precise computerized tests can be debated, also considering that long test sessions hamper the assessment's validity due to mental fatigue and flaws in concentration because of boredom.

Challenge 2: technical limitations hamper the use of computerized experimental paradigms in clinical settings

Most paradigms used in CNS have been developed at universities or research institutes. That means that they have often been programmed using lab software such as Presentation (Neurobehavioral Systems, Inc., Berkeley, CA, USA, www.neurobs.com) or PsychoPy (Peirce, 2007). Software like this often requires full administrator rights of the pc on which the software is running, which in clinical settings is rarely the case as only 'supported' software is allowed to be installed and maintained. Additionally, the use of desktop personal computers is increasingly phased out in clinics in favor of server-based work stations, on which running such software is even more difficult. The information technology (IT) infrastructure in clinics is also often lagging behind in terms of operating systems, internet browsers or plug-in versions, which further complicates the use of customized software. Moreover, the use of touch-sensitive screens

or dedicated button boxes also requires specific hardware. While some of these limitations can be overcome by the use of, for instance, stand-alone laptops dedicated for test administration, data management can be complex as well. Most experimental paradigms, for example, do not issue summary reports that can be used for clinical interpretation, but raw output files at trial level. Data cannot be easily integrated with electronic patient files and data safety cannot always be guaranteed if data have to be physically moved using universal serial bus (USB) drives or sent via unsafe network connections such as wireless local area networking (Wi-Fi), if these are available in the first place. Finally, there is not one standard experimental software environment, but many different software packages that need to be installed and maintained if one wants to use different experimental tasks in the clinic. While these limitations may seem trivial to those not working in clinical settings, it is factors like this that seriously hamper the implementation of experimental computerized paradigms for neuropsychological assessment.

Challenge 3: psychometric properties of CNS paradigms are under-examined or poor

Although many experimental paradigms have been used in numerous studies in the field of CNS, these studies rarely deal with the psychometric properties of the paradigms themselves but focus on the relation between the task's outcome and, for instance, neuroimaging correlates or differences between specific task manipulations. However, in clinical neuropsychology, tests have to be reliable and show good validity (Lezak et al., 2012). Unfortunately, even established experimental paradigms suffer from psychometric shortcomings. For instance, reliability coefficients as low as .21 have been reported for the *N*-back paradigm (Jaeggi, Buschkuehl, Perrig, & Meier, 2010), a widely-used task to study working memory processing. As a result Jaeggi et al. (2010) conclude that the *N*-back paradigm should not be used for measuring individuals differences in working memory performance, as an unreliable test by definition has poor validity. Concerning validity, several critical limitations have been identified by Buelow and Suhr (2009) concerning the IGT, a widely-studied decision-making task examining risk-taking behavior. One such limitation is that performance on the IGT appears to be highly dependent on the participant's personality and mood state. The authors conclude that '… these limitations may affect the IGT's ability to be used effectively as a clinical instrument to judge an individual's performance' (p. 111).

Another concern with CNS tasks is that they are *paradigms* rather than *tests*. That is, a test typically consists of a standard set of stimuli in a fixed order, which has been validated and examined in both normal participants and patient samples, administered in accordance with strict guidelines and instructions. Most CNS paradigms lack this standardization, as the number of trials differs across studies, the type of stimuli can vary (verbal vs visuospatial) as well as the response mode (button-press, response box, touch-sensitive screen). This makes implementation of especially novel paradigms difficult. For instance, in their review, Neguţ, Matu, Sava, and David (2016) concluded with respect to virtual reality assessment in ADHD, while promising, only one task was standardized in such a way that it could be used in neuropsychological assessment of

attention deficits (i.e. the Advanced Virtual Reality Tool for the Assessment of Attention [AULA]; Iriarte, Diaz-Orueta, Cueto, Irazustabarrena, Banterla, & Climent, 2016).

Finally, even when experimental paradigms are available in a standardized format, the lack of available norms may prohibit their use in clinical practice. For instance, freely-available tasks such as the Working Memory Test Battery (containing multiple working-memory paradigms; Stone & Towse, 2015) or the Box Task (a visuospatial working memory paradigm; Kessels & Postma, in press) come with standardized stimulus sets. However, in clinical neuropsychology, clinical decisions have to be made at the $N = 1$ level, which means that confounding factors such as age, gender, socioeconomic status and education level – which are known to affect the performance on most if not all cognitive tasks – should be taken into account. Group means and standard deviations reported in scientific papers rarely allow for these decisions, as groups are small and often highly homogeneous (i.e. psychology undergraduates aged 18–21). Large normative samples of individuals who completed a fixed and standardized version of a given task, in which the effects of the aforementioned confounding variables can be studied and which can be used to transform an individual performance into a standardized outcome measure (e.g. percentile, *T* score) are required.

Challenge 4: CNS paradigms often rely on reaction times rather than accuracy

For measuring the effects of many experimental manipulations in CNS, reaction times are more sensitive than accuracy. Many classic findings are even based on reaction times only, as accuracy measures often produce near-ceiling effects. For instance, in semantic priming experiments, semantically related words presented briefly prior to a target word facilitate decisions about these target words in terms of reaction times (i.e. making a word/non-word decision). In absolute terms, however, such effects can be very small, as semantic priming effects of less than 20 ms have been reported (Bodner & Masson, 2003). Furthermore, the presentation duration of the primes (and targets) in such experiments are also very short (e.g. less than 50-ms presentation duration for a prime stimulus; see Bodner and Masson, 2003).

Importantly, most of these established priming effects have been demonstrated in undergraduate psychology students, but it is unclear whether these generalize to the 'normal population'. That is, reaction-time variability in non-student participants can be higher than the actual effect under study. Specifically effects of aging are important in this respect, as older adults may need 3 to 4 times more time to identify stimuli after a mask than younger adults, resulting in large performance differences (Coyne, 1981). Furthermore, while undergraduate students who may regularly participate in experiments are 'trained' in performing tasks in which reaction times are the outcome measure and consequently respond as fast as possible, naïve and/or older participants may adopt a different response strategy, for example, by favoring accuracy over speed (Forstmann, Tittgemeyer, Wagemakers et al., 2011). Overall reduced processing speed in older adults (Salthouse, 1996), a key target group in clinical neuropsychology, may thus seriously lower the validity and complicate the interpretation of the performance on either speeded stimulus presentation or reaction times as outcome measures.

Challenge 5: unclear ecological validity

Most studies in CNS deal either with the cognitive or neural underpinnings of specific task outcomes. Many of these paradigms have indeed been studied in a variety of patient samples and have often been related to detailed lesion information, specific brain dysfunction, or patterns of brain activation, providing evidence for the task's construct validity. Understanding the underlying neurocognitive mechanisms may also be important in neuropsychological assessment. The extent to which a specific task actually measures the cognitive function or domain under study (*Is the lowered performance on a memory task due to an executive control deficit or a mnemonic processing impairment?*) or is sensitive to specific types of brain injury (*Is the task performance related to prefrontal dysfunction or not?*) determines the task's ability to make a diagnostic prediction. However, in clinical practice it is often more important to establish whether impairments may interfere with return to work or preclude living at home independently, i.e. to make a functional prediction. The latter question is closely related to the concept of ecological validity, specifically a task's veridicality, i.e. the extent in which a performance on a tasks predicts the performance on a criterion related to everyday life. This type of predictive validity has been studied for many traditional neuropsychological tests (e.g. Larson, Kirschner, Bode et al., 2005), but the degree in which outcome measures of paradigms in CNS predict behavioral problems, return to work, the course of decline or treatment response is largely unclear, even in the case of experimental paradigms that have been studied in various patient samples.

Another important aspect of ecological validity concerns verisimilitude. While not a psychometric form of validity in the strict sense, verisimilitude is considered relevant in neuropsychological assessment: do the tasks or tests resemble in any way everyday-life challenges (Chaytor & Schmitter-Edgecombe, 2003)? Even established neuropsychological tests can be criticized in this respect, but tests have been developed with the aim of resembling everyday-life task demands, such as the Rivermead Behavioural Memory Test (Wilson, Greenfield, Clare et al., 2008) or the BADS. Tasks that resemble everyday demands are often more appealing for patients, possibly resulting in optimal motivation to perform as well as possible. In contrast, lengthy computerized paradigms using abstract stimuli and requiring many trials can be perceived as tedious an unmotivating, which may affect the performance validity.

Challenge 6: technical progress affects development and availability

Since many – if not all – cognitive paradigms used in modern-day CNS are administered using computers, technological progress dictates the rate of development and maintenance of computer software to be used. Computing power has been argued to double every year since the 1970s (a phenomenon known as Moore's Law), resulting in continuous updates and upgrades of operating systems and hardware components. There are, however, many examples of computerized test batteries that are being used in the clinic – albeit more as trial outcome measures than as part of routine testing – and some of these also include paradigms originating from CNS (e.g. CNS Vital Signs, Cogstate, or the Cambridge Neuropsychological Test Automated Battery

[CANTAB]; see Parsey & Schmitter-Edgecombe, 2013, for a review). Also, many widely-used attention tests or batteries are successful examples of originally experimental computerized paradigms that are being used in clinical practice (e.g. Conners' Continuous Performance Test, Conners & Staff, 2015, or the Test of Attentional Performance, Zimmermann & Fimm, 2002).

However, changes in software or hardware may also affect these tasks' validity, as variables like stimulus size and the accuracy of response time measurement may change with every update. In theory, this would also require re-norming and re-validating of these tasks. However, the cost and time needed associated with such an enterprise prohibit this in practice, especially in the case of open-source or public domain software, which is not supported by a business model to earn back investments through revenues. Also, by the time a new computer version of test has been fully validated and normed, it is likely that further updates have been deemed necessary. This problem does not affect traditional paper-and-pencil tasks: the stimulus words of the Hopkins Verbal Learning Test can be held constant for decades and the stimulus sheet of the TMT has remained largely unaltered since its first use in neuropsychological practice.

Furthermore, more recent development with respect to cloud computing and the use of tablets or even smart phones rather than desktop computers provide further challenges. From the user perspective, data-safety and privacy rules and regulations have already been discussed (see Challenge 2), but new developments like these are also challenging for test developers. Tablets may on the one hand be useful tools for facilitating computerized assessments, as they resemble the paper-and-pencil format more than a computer-and-mouse setup (see e.g. Fellows, Dahmen, Cook, & Schmitter-Edgecombe, 2017, on a tablet version of the TMT), on the other hand the actual user interface differs greatly from actual paper-and-pencil administration. Studies examining the equivalence of computer versions with the traditional paper-and-pencil form or re-norming may thus be required.

Precision neuropsychology: the need for novel paradigms

Although it may be challenging to use novel paradigms developed and studied by cognitive neuroscientists in clinical neuropsychology, classic paper-and-pencil tasks are by no means always superior. Even the validity of widely used tasks such as the Stroop Color-Word Test has been criticized (Eling, 2018), normative samples are sometimes small and important predictors such as education are not always taken into account (e.g. for the BADS, see Strauss et al., 2006), and test-retest reliability can be severely affected by practice effects, even when using parallel versions (e.g. for Rey's Verbal Learning Test; Van der Elst, Van Boxtel, Van Breukelen, Jolles, 2008). Clearly, there is need for novel paradigms that are theory-driven on the one hand, but on the other hand can be related to clinical outcomes.

Many existing paradigms from CNS could be adapted for use in clinical assessments. Also, novel approaches such as the use of virtual reality, or ecological momentary assessments (experience sampling), in which data are collected multiple

times a day by the patient in their home environment (Parsey & Schmitter-Edgecombe, 2013), using for example tablets or wearables, have the potential to innovate neuropsychological assessments. Such innovations may facilitate measurements and consequent interpretations that are not possible using classic paper-and-pencil paradigms or computerized versions thereof. Cognitive neuroscientist need to collaborate with clinical neuropsychologists though, in order to achieve this in such a way that these tasks result in answers to clinically relevant questions. Moreover, we should not necessarily shy away from the paper-and-pencil format of assessing patients, as reliable experimental findings should be robust enough to be detected by paper-and-pencil measures (e.g. the Stroop effect by timing the response with a stopwatch, amnesia by measuring recall of pictures shown on cards or words read aloud, and executive dysfunction by perseveration on tasks requiring task switching or set shifting).

Overcoming the challenges: conclusion

The six challenges addressed in this paper need to be overcome by clinical neuropsychologists in close collaboration with colleagues from the field of CNS. For CNS paradigms to be used in clinical assessment, (1) paradigms should be reduced to the essentials, focusing only on reliable and clinically meaningful outcome variables, (2) tasks need to be platform independent, preferably using internet-based environments that can run on any work station, pc or tablet, (3) more research (and funding) is needed on the psychometric properties of experimental paradigms, (4) experimental tasks should be adapted for use in the clinic in such a way that accuracy becomes more important than reaction time, (5) computer tasks could be made more attractive by gamification of paradigms, also enabling experience sampling, and everyday outcome measures should be included in research on ecological validity, and (6) open-access platforms should be made self-sustainable in order to facilitate the long-term availability of tasks.

Note

1. Note that obviously, a neuropsychological assessment is not limited to administering tests and interpreting their outcome in a quantitative way, but also involves a clinical interview with the patient and his/her spouses, medical history taking, behavioral observation and the interpretation of non-cognitive factors, such as personality, wellbeing and psychological symptoms, which should all be integrated with the results of neuropsychological tests.

Acknowledgments

This article is based on a presentation given by the author at the 2016 Mid-Year Meeting of the International Neuropsychological Society held in London, UK.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Roy P. C. Kessels http://orcid.org/0000-0001-9500-9793

References

Bauer, R.M., Iverson, G.L., Cernich, A.N., Binder, L.M., Ruff, R.M., & Naugle, R.I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 177–196. doi: 10.1080/13854046.2012.663001.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159–162. doi: 10.1016/j.tics.2005.02.002

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*, 383–405. doi: 10.1076/jcen.24.3.383.981

Bodner, G.E., & Masson, M.E. (2003). Beyond spreading activation: An influence of relatedness proportion on masked semantic priming. *Psychonomic Bulletin and Review*, *10*(3), 645–652. doi: 10.3758/BF03196527

Buelow, M.T., & Suhr, J.A. (2009). Construct validity of the Iowa Gambling Task. *Neuropsychology Review*, *19*(1), 102–14. doi: 10.1007/s11065-009-9083-4.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–97. doi: 10.1023/B:NERV.0000009483.91468.fb

Conners, C.K., & Staff, M. (2015). *Conners' continuous performance test* (3rd ed.) (Conners CPT 3). North Tonwanda, NY: Multi-Health Systems Inc.

Coyne, A. C. (1981). Age differences and practice in forward visual masking. *Journal of Gerontology*, *36*, 730–732. doi: 10.1093/geronj/36.6.730

Eling, P. (2013).Op het spoor van een 'multifunctionele' test: Over de geschiedenis van de Trail Making Test [Searching for a 'multifunctional' test: On the history of the Trail Making Test]. *Tijdschrift voor Neuropsychologi.e*. *8*, 53–58.

Eling, P. (2018). Waarom gebruikt u eigenlijk de Stroop? [Why do we actually use the Stroop?]. *Tijdschrift voor Neuropsychologi.e*. *13*, 68–72.

Fellows, R.P., Dahmen, J., Cook, D., & Schmitter-Edgecombe, M. (2017). Multicomponent analysis of a digital Trail Making Test. *The Clinical Neuropsychologist*, *31*(1), 154–167. doi: 10.1080/13854046.2016.1238510.

Forstmann, B.U., Tittgemeyer, M., Wagenmakers, E.J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience*, *31*(47), 17242–17249. doi: 10.1523/JNEUROSCI.0309-11.2011.

Gazzaniga, M.S., Ivry, R., & Mangun, G.R. (2009). *Cognitive neuroscience: The biology of the mind* (3rd ed.). New York, NY: W. W. Norton.

Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., & Climent, G. (2016). AULA – Advanced Virtual Reality Tool for the Assessment of Attention: Normative Study in Spain. *Journal of Attention Disorders*, *20*, 542–68. doi: 10.1177/1087054712465335

Jacobs, J. (1887). Experiments on "prehension." *Mind*, *12*, 75–79. doi: 10.1093/mind/os-12.45.75

Jaeggi, S.M., Buschkuehl, M., Perrig, W.J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394–412. doi: 10.1080/09658211003702171

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358. doi: 10.1037/h0043688

Kessels, R.P.C., & Postma, A. (in press). The Box Task: A tool to design experiments for assessing visuospatial working memory. *Behavior Research Methods*. doi: 10.3758/s13428-017-0966-7

Larson, E., Kirschner, K., Bode, R., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the Repeatable Battery for the Assessment of Neuropsychological Status in the evaluation of stroke patients. *Journal of Clinical and Experimental Neuropsychology*, *27*(1), 16–32. doi: 10.1080/138033990513564

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York, NY: Oxford University Press.

Maes, J.H.R, Eling, P.A.T.M., & Reelick, M.F., & Kessels, R.P.C. (2011). Assessing executive functioning: on the validity, reliability, and sensitivity of a click/point random number generation task in healthy adults and patients with cognitive decline. *Journal of Clinical and Experimental Neuropsychology*, *33*, 366–378. doi: 10.1080/13803395.2010.524149

Neguţ, A., Matu, S. A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: A meta-analytic review. *The Clinical Neuropsychologist*, *30*(2), 165–184. doi: 10.1080/13854046.2016.1144793

Nissen, M.J., & Bullemer, P. (1987) Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32. doi: 10.1016/0010-0285(87)90002-8

Oosterman, J.M., Vogels, R.L., van Harten, B., Gouw, A.A., Poggesi, A., Scheltens, P., Kessels, R.P.C., & Scherder, E.J. (2010). Assessing mental flexibility: Neuroanatomical and neuropsychological correlates of the Trail Making Test in elderly people. *The Clinical Neuropsychologist*, *24*, 203–219. doi: 10.1080/13854040903482848

Parsey, C.M., & Schmitter-Edgecombe, M. (2013). Application of technology in neuropsychological assessment. *The Clinical Neuropsychologist*, *27*(8), 1328–1361. doi: 10.1080/13854046.2013.834971

Peirce, J.W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique: Les problems. *Archives de Psychologi.e*. *28*, 215–285.

Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. doi: 10.1037/0033-295X.103.3.403

Stone, J.M. & Towse, J.N., (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software*, *3*(1), e5. doi: http://doi.org/10.5334/jors.br

Strauss, E., Sherman, E. M. S. & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi: 10.1037/h0054651

Van der Elst, W., Van Boxtel, M.P., Van Breukelen, G.J., & Jolles, J. (2008). Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's Verbal Learning Test, and the Letter Digit Substitution Test: The regression-based change approach. *Journal of the International Neuropsychological Society*, *14*(1), 71–80. doi: 10.1017/S1355617708080028

Wilson, B.A., Greenfield, E., Clare, L., Baddeley, A., Cockburn, J., Watson, P., & Nannery, R. (2008). *The rivermead behavioural memory test*. 3rd edn. London: Pearson Assessment.

Zimmermann, P., & Fimm, B. (2002). A test battery for attentional performance. In M. Leclercq & P. Zimmermann (Eds.), *Applied neuropsychology of attention: Theory, diagnosis and rehabilitation* (pp. 110–151). London: Psychology Press.