Published in final edited form as:

Estud Psicol. 2014 July 25; 35(2): 426–437. doi:10.1080/02109395.2014.922264.

What standardized tests ignore when assessing individuals with neurodevelopmental disorders

Marcela Tenorio^a, Ruth Campos^b, and Annette Karmiloff-Smith^c

^aPontificia Universidad Católica de Chile

^bUniversidad Autónoma de Madrid

^cBirkbeck College, University of London

Abstract

In this article we critique the use of traditional standardized tests for the cognitive assessment of children with neurodevelopmental disorders. Limitations stem from the lack of integrating (a) results from research into the psychological functioning of these populations, and (b) the main arguments underlying models of human development. We identify four secondary issues in this discussion: (1) these instruments cannot be used with children who have particularly low cognitive functioning; (2) little or no variance in the scores obtained by individuals with neurodevelopmental disorders, because all are at floor, prevent adequate interpretations; (3) measurements do not provide information useful for the design of intervention strategies; and (4) different cognitive and/or neural processes may underlie behavioural scores 'in the normal range'. Rethinking traditional assessment methods in favour of technologically-mediated games yields new cognitive assessment possibilities.

Keywords

test standardization; evaluation; neurodevelopmental disorders; education

Traditional tests are standardized on typically developing populations and, as a result, they tend to ignore crucial factors in the cognitive assessment of individuals with neurodevelopmental disorders (NDDs). Among such factors that are typical of NDDs are: fatigue, boredom, low self-esteem, difficulties in motor control, difficulties in understanding task instructions and stress.

Children with NDDs are often subject to long and tedious cognitive assessment sessions, where tests (mostly paper and pencil tests, with high demands on fine motor coordination) are administered that aim to establish a performance profile (Miller, 2013). The results of the assessments often only capture a static image of the participant, located at a single point in time and space that fails to measure developmental change. In other words, such measurements do not provide information on the individual's developmental process nor do

they allow extrapolation in terms of future developmental change. Due to their adherence to the instruction manuals, it is common for clinicians and researchers to forget to question the client or the people close to them about their development path, or to find other ways of phrasing their questions.

Recent research has identified several relevant considerations: (1) working with children involves taking an ontogenetic perspective (Karmiloff-Smith, 1998, 2010); (2) a static view of cognitive development does not yield information needed to design interventions; (3) behavioural data do not always provide information about underlying cognitive and neural processes (i.e. data that fall 'within a normal range' often camouflage neural differences between typical and atypical development); and (4) the study of infant cognition requires special models, where development is a fundamental variable.

This paper aims to demonstrate why traditional tests are under crossfire because of problems that are inherent to their format (Rosas & Ceric, 2011) but also because of their most important theoretical limitation: the impossibility of tracing developmental trajectories with the measurements that they provide. Beyond this central issue, we also point to some minor difficulties and discuss them with reference to empirical data.

We have identified two problems related to the tests themselves and two other problems related to the use professionals make of the tests. Regarding the first: (1) by virtue of having been designed and standardized using samples who are typically developing, their use with atypically developing children is inappropriate, and (2) within NDDs there is a wide individual variation, but because many individuals are at floor on the tests, these individual differences are camouflaged. Regarding the problems that arise from using the tests: (3) they are often used for educational decisions, despite the fact that they do not yield any information on intervention strategies and most importantly, (4) behavioural scores 'in the normal range' tend to be erroneously considered as 'preserved' functioning, when often the cognitive and neural processes underpinning them are different from people with typical development.

If the client provides no data, this is nonetheless considered a valid datum

One of the main problems with the use of traditional standardized tests is that they may be impossible to administer to individuals with NDDs, even by a very experienced clinician, as the following study illustrates.

Thirty-eight children ($M = 7.3$ years) with genetic syndromes associated with intellectual disabilities (Down Syndrome [DS], Williams Syndrome [WS] and Fragile X Syndrome [FXS]), residents of urban areas in Chile, were invited to participate in several research projects exploring the feasibility of using digital games for cognitive assessment. As part of the process, a traditional standardized test was also administered, for which the third edition and Chilean version of the Wechsler Intelligence Scale for Children, (WISC-IIIv.ch., Ramírez & Rosas, 2007) was selected, because it is the only tool available at a national level with local standardization data.

Scheduled sessions were suspended many times for various reasons. First, the analysis of the videos of the sessions revealed that most of these children used non-verbal communication. Oral instructions such as 'please give me your best meaning for each of the words that I am going to say to you' could not be accompanied by signed manual instructions. As a result, we observed behavioural patterns that negatively impacted performance (e.g., echolalia and echopraxia as well as severe difficulties in maintaining attention on the task and on the evaluator's instructions).

The difficulties of interpreting floor effects

During cognitive assessments, particularly with children with NDDs, it is natural that an emotional bond is created between the evaluator and person being evaluated, which cannot be ignored. Yet, according to Manual instructions, feedback is limited and often the only phrases allowed are 'continue' or 'try again'. This situation is particularly disconcerting for children with NDDs, who often expect much more support while performing their tasks. In turn, the situation is also frustrating for the evaluator as the child's failure becomes imminent and restrictions imposed by test instructions do not allow any modifications to the administration procedure. In such circumstances, the child usually gives very few correct answers, so the assessment is suspended and the evaluation ends with little information. In the Wechsler Scales of Intelligence or the Kaufman Assessment Battery for Children, among others, the minimum IQ values that can be scored are the floor effects of performance on several subtests, yielding a flat profile, a ratio of no more than 60 points and no information about differences across specific domains (and of course, no data on the course of development).

This lack of intra- and inter-individual variation has made it very difficult to differentiate between groups with different NDDs. For example, in terms of literacy, given their low performance on traditional tests, it was assumed that children with DS and WS could not acquire reading skills, and that there were no differences between the syndromes, because no examination was carried out into the causes of their deficits or error patterns. However, thanks to data from basic research it has become clear that there are marked and early cross-syndrome differences in key precursors to literacy such as phonological awareness (Steele, Scerif, Cornish, & Karmiloff-Smith, 2013). It is crucial that such differences are incorporated into the reading assessment battery design: that basic research is translated into relevant diagnostic tasks for NDDs.

In the case of mathematics, the floor effects observed in traditional tests once again make the performance of infants and toddlers with different NDDs seem very similar. However, a deeper cognitive exploration reveals different cross-syndrome strengths and weaknesses. For example, infants with WS are similar to their typically developing peers when differentiating small numbers (e.g., 2 vs. 3), but not when larger amounts (magnitudes like 8 vs. 16) are involved, while infants with DS show the opposite pattern (Karmiloff-Smith et al., 2012; Van Herwegen, Ansari, Xu, & Karmiloff-Smith, 2008).

Floor effects on standardized tests create the illusion of homogeneity and make the design of interventions targeting the particular needs of each NDD and each individual child difficult.

Using standardized tests in the design of curricular supports

We now consider limitations of the use of traditional tests with respect to their implications for decisions about education. If results are all at floor, why bother to carry out a cognitive assessment on a school-aged child if the results will not significantly contribute to the design and implementation of a curriculum that will most benefit him? Psychometric information provided by traditional tests may provide very general scores, but with regard to aiding in decision-making about educational methods and clinical interventions, specific performance data are critically needed. We know, for example, that individuals with WS have problems in eye movement planning and attentional shifts, which implies atypical patterns in scanning quantities which affect large number discrimination, while individuals with DS have difficulties in sustained attention on individual objects which affect small number discrimination (Ansari & Karmiloff-Smith, 2002). Curricula designed for school aids and specialized interventions should consider these cross-syndrome differences, but usually a one-fit-all method is used for all children with special educational needs.

Along with standard batteries and the use of specific cognitive tasks, other aspects should be included in a full assessment, such as a person's social or autonomous skills, their interests and motivations, and the possibilities provided by their environment to support their development. The inclusion of these aspects, in addition to conducting assessments in more natural contexts, would help resolve the issue of lack of ecological validity of standardized tests.

What is concealed under scores 'in the normal range'

Despite the widespread belief that any score in the normal range on a standardized test can be taken to reflect neurotypical functioning, there is telling evidence pointing to the inadequacy of analyses based solely on behavioural results. For example, Tomalski et al. (2013) studied the impact of differences in socio-economic levels on the brain development of 6- to 9-month-old infants who lived in areas of social deprivation in East London. Infants from families with high or low socio-economic levels obtained scores 'in the normal range' on the Bayley Scales of Infant Development. However, when EEG measurements were taken, a difference emerged in the gamma-band activity in the frontal cortex between the two socio-economic groups. Those from low SES had significantly reduced activity, suggesting less capacity for information integration. These findings suggest that family income level influences brain development as early as six months of age and, despite behavioural scores that fall 'in the normal range', children without genetic mutations but who have grown up in lower socio-economic environments exhibit atypical neural development.

Differences between behavioural levels and their neural substrates are even more obvious in children with NDDs. In a study with adolescents and adults with autism spectrum disorders, memory tasks yielded no differences with people with typical development at the behavioural level, but electrophysiological measurements of brain activity reflected atypical temporal patterns in those with autism (Massand, Bowler, Mottron, Hosen, & Jemel, 2013). Similarly, scores for individuals with WS on standardized face processing tests (e.g., Benton

or Rivermead) fall 'in the normal range', but again the underlying neural patterns reveal atypical brain activity (Karmiloff-Smith et al., 2004). In short, many standardized tests are not sensitive enough to reveal atypical development paths or causes of behavioural deficits, resulting in a loss of important information.

The transformation of cognitive assessment

Obviously not all children coming in for clinical assessments can be subjected to neural measurements typical of basic research. So how can we improve behavioural assessments so that they tap underlying cognitive processes? Here we make some basic recommendations that can be incorporated into the new generation of tests. First, traditional assessments usually engage explicit cognitive systems, that is, metacognitive processes, where individuals must explicitly reflect on the questions and their answers. Since those with NDDs frequently have a deficit in metacognitive skills, this presupposes something which is sometimes impossible for them. It would thus be interesting to explore the use of activities whose design facilitates the engagement of *implicit* processing systems, where individuals with NDDs frequently show better performance (Don, Schellenberg, Reber, DiGirolamo, & Wang, 2003).

In addition, tasks should be transformed into games, to increase motivation and to facilitate capturing and maintaining the attention of those with NDDs. Games are a creative way to turn difficult problems into accessible tasks that such individuals can understand (Rosas & Ceric, 2011). If a technological platform is added to the game context, we have created a bridge between evaluator and the person being evaluated, one that motivates, supports activities, facilitates engagement and maintains attention.

Recognizing the limitations and dangers of traditional tests, we raise one final consideration: their use and the scores they yield cannot be considered mandatory, either as a constraint for receiving educational benefits or in the design of teaching aids. Prior to creating legislation in this regard, it is important to support research agendas that will enable a better understanding of the range of psychological functioning in NDD populations.

Acknowledgements

MT and AK-S are grateful to the British Academy's International Mobility Project which supported this work. MT thanks the Government of Chile FONDECYT post-doctoral project 3130612 and FONDECYT regular grant 1110975. RCG is grateful for funding from the Spanish Ministry of Economy and Competitiveness PSI 2010-21457 project. AK-S also thanks the Wellcome Trust Strategic Award (Grant 098330/Z/12/Z), the Waterloo Foundation and Autour de Williams.

References

- Ansari D, Karmiloff-Smith A. Atypical trajectories of number development: A neuroconstructivist perspective. *Trends in Cognitive Sciences*. 2002; 6:511–516. doi:10.1016/S1364-6613(02)02040-5. [PubMed: 12475711]
- Don AJ, Schellenberg EG, Reber AS, DiGirolamo KM, Wang PP. Implicit learning in children and adults with Williams syndrome. *Developmental Neuropsychology*. Special Issue: Williams Syndrome. 2003; 23:201–225. doi:10.1080/87565641.2003.9651892. [PubMed: 12730025]
- Karmiloff-Smith A. Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*. 1998; 2:389–398. doi:10.1016/S1364-6613(98)01230-3. [PubMed: 21227254]

- Karmiloff-Smith A. Neuroimaging of the developing brain: Taking 'developing' seriously. *Human Brain Mapping*. 2010; 31:934–941. doi:10.1002/hbm.21074. [PubMed: 20496384]
- Karmiloff-Smith A, D'Souza D, Dekker TM, Van Herwegen J, Xu F, Rodic M, Ansari D. Genetic and environmental vulnerabilities in children with neurodevelopmental disorders. *Proceedings of the National Academy of Sciences*. 2012; 109:17261–17265. doi:10.1073/pnas.1121087109. [PubMed: 23045661]
- Karmiloff-Smith A, Thomas M, Annaz D, Humphreys K, Ewing S, Brace N, Campbell R. Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry*. 2004; 45:1258–1274. doi:10.1111/j.1469-7610.2004.00322.x. [PubMed: 15335346]
- Massand E, Bowler DM, Mottron L, Hosen A, Jemel B. ERP correlates of recognition memory in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2013; 43:2038–2047. doi:10.1007/s10803-012-1755-x. [PubMed: 23307419]
- Miller, D. *Essentials of school neuropsychological assessment*. Wiley & Sons; New York, NY: 2013.
- Ramírez V, Rosas R. Estandarización del WISC-III en Chile: Descripción del Test, Estructura Factorial y Consistencia Interna de las Escalas. *Psyche*. 2007; 16(1):91–109. doi:10.4067/S0718-22282007000100008.
- Rosas, R.; Ceric, F. ¿ Pruebas tradicionales o evaluación invisible a través del juego?: Nuevas fronteras de la evaluación cognitiva. 2011. Unpublished manuscript
- Steele A, Scerif G, Cornish K, Karmiloff-Smith A. Learning to read in Williams syndrome and Down syndrome: syndrome-specific precursors and developmental trajectories. *The Journal of Child Psychology and Psychiatry*. 2013; 54:754–762. doi:10.1111/jcpp.12070. [PubMed: 23718731]
- Tomalski P, Moore DG, Ribeiro H, Axelsson EL, Murphy E, Karmiloff-Smith A, Kushnerenko E. Socioeconomic status and functional brain development - associations in early infancy. *Developmental Science*. 2013; 16:676–687. doi:10.1111/desc.12079. [PubMed: 24033573]
- Van Herwegen J, Ansari D, Xu F, Karmiloff-Smith A. Small and large number processing in infants and toddlers with Williams syndrome. *Developmental Science*. 2008; 11:637–643. doi:10.1111/j.1467-7687.2008.00711.x. [PubMed: 18801117]